AWARD NUMBER: W81XWH-15-C-0141

TITLE: An Assessment Tool to Detect Unique Characteristics of Cognitive Deficiency

PRINCIPAL INVESTIGATOR: Corinna Lathan, Ph.D.

CONTRACTING ORGANIZATION: AnthroTronix, Inc.
                                                         Silver Spring, MD 20910

REPORT DATE: June 2017

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                           Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
June 2017	Final	30 Sep 2015 - 29 JUN 2017

4. TITLE AND SUBTITLE
An Assessment Tool to Detect Unique Characteristics of Cognitive Deficiency

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-C-0141

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Corinna Lathan, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AnthroTronix, Inc.

Silver Spring, MD 20910

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This grant prepares DANA (Defense Automated Neurobehavioral Assessment), the next-generation neurocognitive assessment tool (NCAT), for transition into operational military use. DANA is a clinical decision support tool developed for and funded by the Department of Defense (DoD) for use in field and clinical settings.

The project proposed here seeks to 1) analyze our existing data that was collected during prior efforts, in order to discover unique characteristics of psychological impairment and cognitive deficiency that results from either physical trauma, emotional distress or a combination of both factors, which we call DANA Cognitive Deficiency Signature Assessment Tool (DANA CODE SAT), 2) update technical features, and 3) transition DANA to military acquisition customers and programs.

15. SUBJECT TERMS-

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	202	

Table of Contents

INTRODUCTION

"An Assessment Tool to Detect Unique Characteristics of Cognitive Deficiency" is an 18-month effort to prepare DANA (Defense Automated Neurobehavioral Assessment) as the next-generation neurocognitive assessment tool (NCAT) for operational military use. DANA is a clinical decision support tool developed for and funded by the U.S. Department of Defense (DoD) for use in the field. The goal of DANA is to assist first- and second-line providers in the field in determining the type of impairment and level of functioning, close to the time of an incident, as well as medical and clinical providers at military treatment facilities.

This project sought to 1) analyze our existing data that were collected during prior efforts, in order to discover unique characteristics of psychological impairment and cognitive deficiency that results from either physical trauma, emotional distress or a combination of both factors, which we call the DANA Cognitive Deficiency Signature Assessment Tool (DANA CODE SAT), 2) update technical features, and 3) transition DANA to military acquisition customers and programs.

KEYWORDS
Neurocognitive, Cognitive Deficiency, DANA, NCAT, Clinical Support Tool

ACCOMPLISHMENTS

The major goals and objectives of this project were the following:

Objective 1: Customized Battery Creation
Develop a drag-and-drop user interface (UI) that will enable users to create custom test batteries in real-time. The drag-and-drop UI will provide users with sub-test selection, the ability to order the sub-tests in whatever manner they wish, and the ability to modify the sub-test parameters (e.g., number of trials in SRT). For the customized test batteries to work in real-time we will build on our existing DANA code; this involves recoding the existing DANA code and developing code for the new drag-and-drop UI feature.

Objective 2: DANA DDM Opt-Out
We will remove the use of the DANA Data Manager (DDM) data management application and switch to a secure cloud-based data management system that enables DANA data to be continuously uploaded to the cloud. We will develop a desktop user interface and dashboard to access the data stored on the cloud in a SQL database, as well as develop plug-ins that will enable DANA data export to standard analysis tool formats (e.g., MATLAB, SPSS, and SAS).

To fully remove the DDM, we will integrate our cloud support into the DANA codebase. We will integrate cloud authentication into DANA to enable users to store their data on the cloud and access the data from the DANA desktop / mobile login. We will create a DANA desktop and mobile login so that users may be able to access the data securely from any device or computer; this will also enable exporting of the data. The dashboard will be password-protected. A general dashboard app will provide metadata: patient names, dates and name of tests (test results are directly accessed via SQL inquiries from the appropriate analysis software - MATLAB, SPSS, etc.).

The DANA data will be stored in a cloud-based secure database*. For the data to sync to the database we will develop proper encryption for wireless transmission. We will integrate encryption with keys (the difference being that the data is passed wirelessly and continuously where previously it passed via wired USB).

To export the DANA data for analysis we will develop a dashboard for the cloud database that enables exporting to various data analysis software (e.g., MATLAB, SPSS, SAS); additionally, we will develop data analysis plug-ins to enable users to export DANA data automatically from the dashboard interface into these programs.

*Hosting the cloud database and web portal on an internet-enabled server is not a part of the deliverable for this contract. Instructions for setting up hosting are included with the cloud database and web portal code repositories (README.md text files).

Objective 3: Analyze and characterize existing datasets, and process-develop the CODE SAT algorithm
Assemble and categorize, by levels of psychological distress (none, moderate, severe), our existing cognitive data in order to understand the implications of psychological factors on cognitive performance. We will analyze these subset categories of information using applied mathematics techniques so as to uncover characteristic patterns of response in normal and in impaired subjects. We will identify these characteristic patterns within and across categories of response so as to develop an algorithm capable of detecting them on the test-taking device, within DANA.

Most DANA subtests require an input response to a presented stimulus and DANA records the time from when the stimulus was presented to the moment of reaction. Each test progresses through a set number of trials, and it is within the distribution of trial-to-trial reaction times that we will uncover the characteristics of each subject category of response. We will first employ the techniques of clustering analysis and time series analysis to elucidate the trial-to-trial variability of responses, by category and by subtest, and from these patterns we will construct an algorithm capable of identifying limits of normality for cognitive performance.

Information garnered from the analytics effort will allow us to determine a set of parameters under which elucidated characteristics of unimpaired response patterns will be identified post-administration from a subject's trial-to-trial test responses.

Objective 4: Validate the CODE SAT algorithm and implement it into the DANA codebase
We will validate the algorithm against previously collected cognitive performance data from subjects with a known physical trauma (mTBI) with and without psychological distress, and from subjects with no physical impairment, but with known psychological distresses, and we will assess the performance and limits of the algorithm in identifying a difference from impaired and unimpaired signatures of response.

With a validated algorithm, we will adapt it for the DANA codebase and implement it into the functionality of DANA.

Final Report (W81XWH-15-C-0141)

Objective 5: Transition DANA for Acquisition
We will collaborate with key stakeholders, including the Neurocognitive Assessment Branch Rehabilitation and Reintegration Division, HQDA, Office of the Surgeon General of the Army; Defense and Veterans Brain Injury Center, AMEDD, and other agencies. We will provide briefings to military leadership, Defense Centers of Excellence, and present findings for this project at scientific meetings, as well as publish findings in Military Medicine and other peer-reviewed journals.

We will develop and deliver a complete proposal to the government for deploying and supporting DANA.

The existing DANA User Guide will be modified for optimal operational use. Data from the previous objectives will be evaluated and inform user guide updates. Menu and reporting options will most likely change and these will be implemented in the updated user guide.

What was accomplished under these goals?

TECHNICAL DEVELOPMENT

Objective 1: Customized Battery Creation. Status: Completed

**Task 1: Develop drag-and-drop interface.**
To accomplish this task, there were numerous key technical changes that needed to be implemented first. In order to develop a drag and drop interface, we had to restructure the original DANA military app's user interface (UI). This restructuring involved devising a new concept of user interaction with the app and then designing new app screens and elements (including new graphics), following best practices per the Android design guidelines.

As development progressed, usability assessments were completed iteratively to uncover any quality assurance issues as well as errors that could be caused by a user. Based on usability testing feedback, we modified (a) the method for deleting subtests from a custom battery, and (b) the existing icon for reordering subtests within a custom battery to make these operations more intuitive.

Below is a comparison of the changes we made over the course of this contract to the app functionality and UI.

Final Report (W81XWH-15-C-0141)

Examples of original DANA Military app graphics and UI prior to changes:

Left-to-right: Login, Subject creation / selection, Screening selection
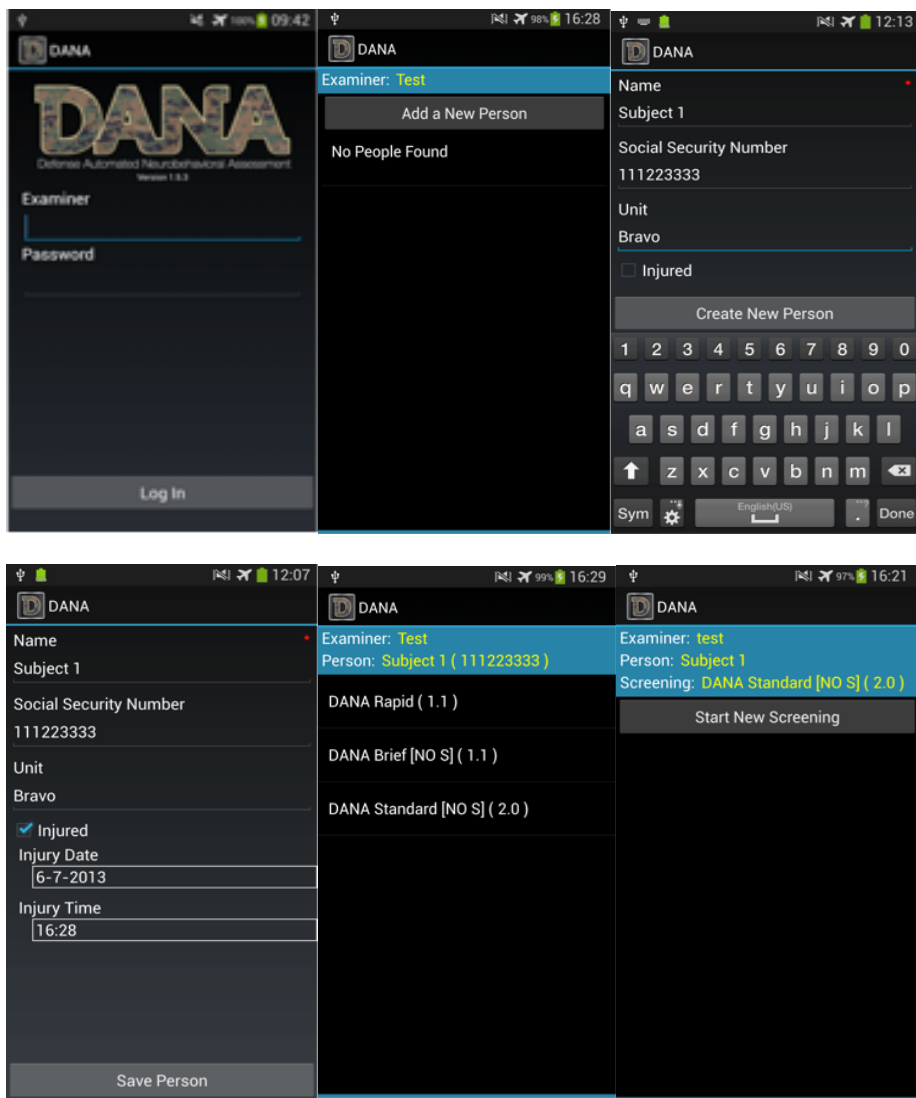
Examples of new DANA Military app graphics and UI after changes:

Changes included the ability to edit existing tests and test batteries, create custom tests and test batteries, as well as the introduction of a drag and drop interface with test icons.

Left-to-right: Login, Subject creation / selection, Test battery customization
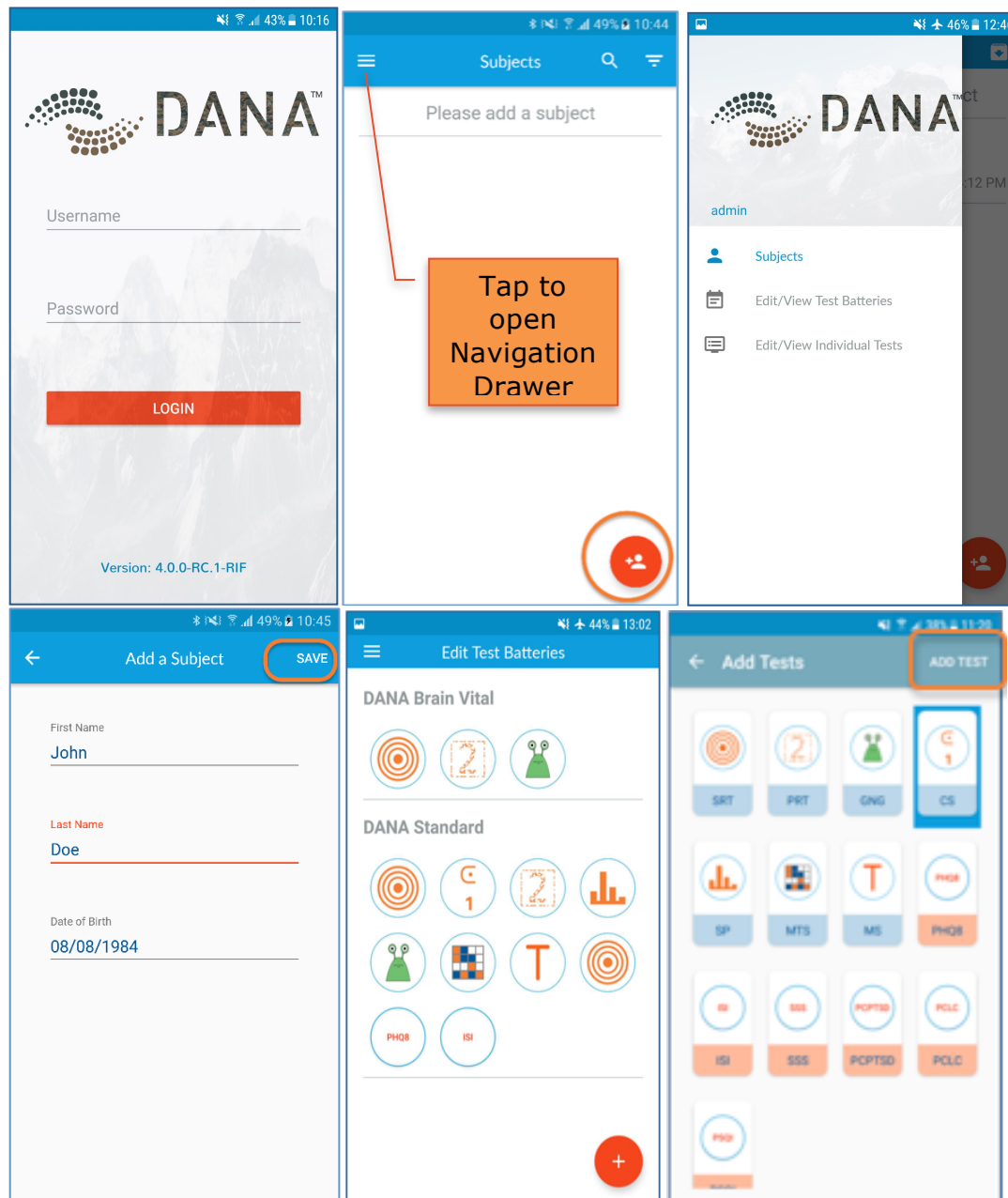
Left-to-right: Test battery customization (continued), Individual test customization
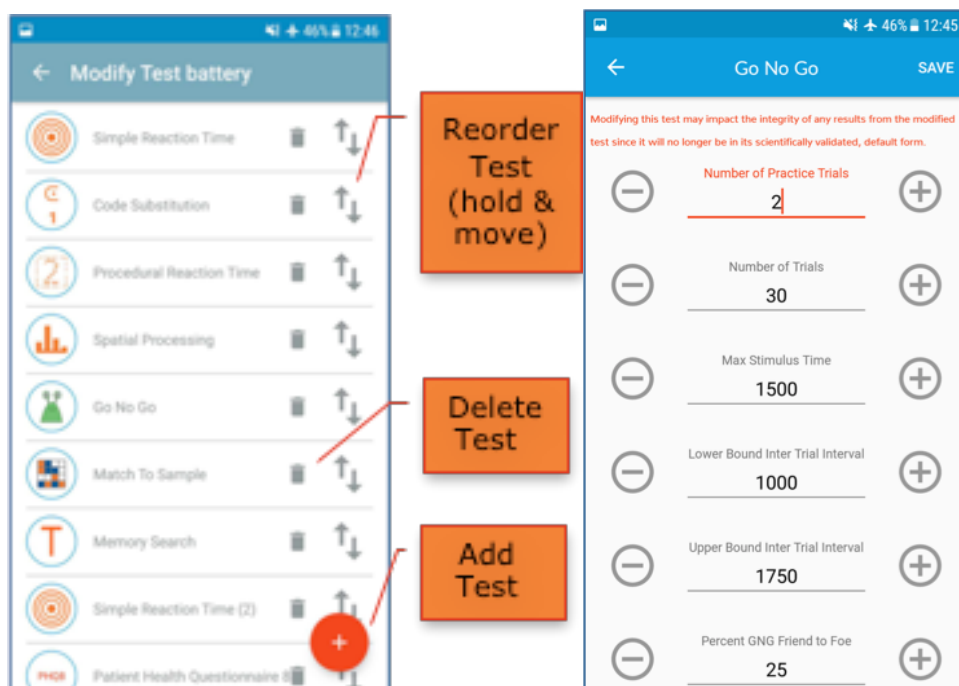
Examples of original DANA military in-app results prior to changes:

Left-to-right: Select completed screening, Summary report, Navigation to detailed report
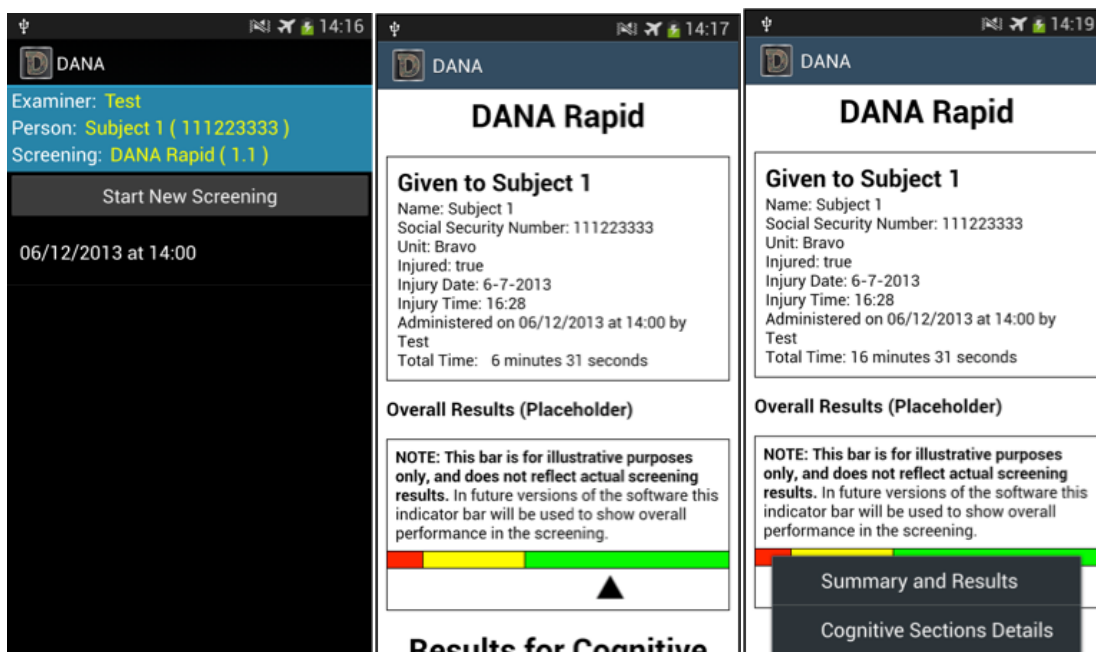
Final Report (W81XWH-15-C-0141)
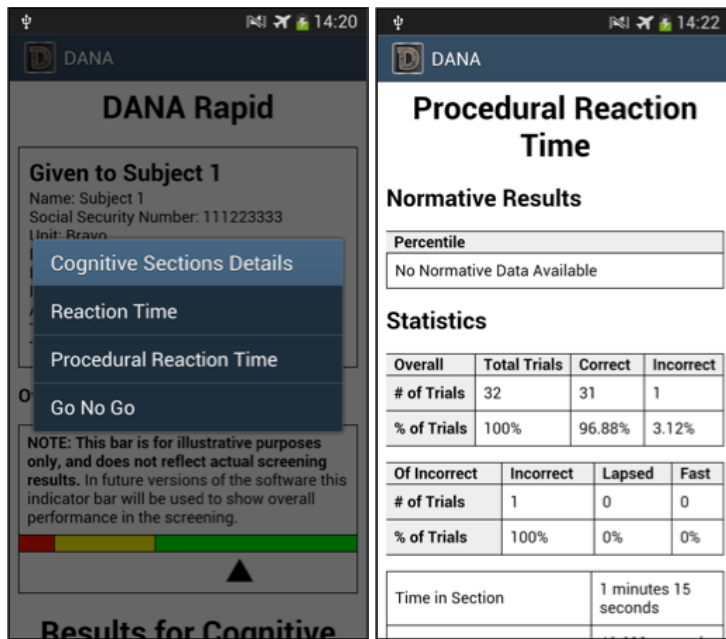
Examples of new DANA military in-app results after changes:

Changes included adding the ability to view graphs of summary scores over time; marking of test scores as acceptable, unacceptable, or incomplete; and updated graphics and navigation.

Left-to-right: Login, View results navigation, Results screens (Summary, Graph, Raw Data)

Final Report (W81XWH-15-C-0141)

On small screens: tap here to toggle to Mean Reaction Time & % Correct

Use the drop down menu to choose which individual test's results to view. Use the bottom tabs to choose graph.

Task 2: Recode DANA for customized batteries.

We modified DANA to allow users to create custom test batteries and cognitive tests. Additional screens were added to the app to guide the user through each process. When customizing a test battery, the user can add or remove any number of the individual tests or surveys and save the new test battery with a unique name. The new test battery will then be available as a screening option within the app. When customizing a cognitive test, the user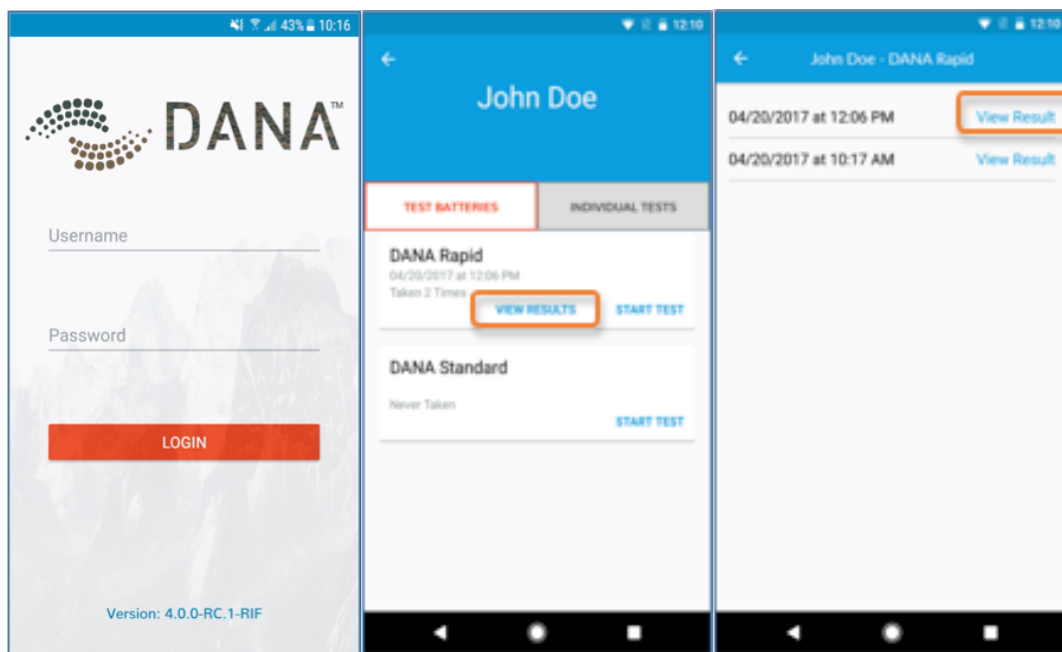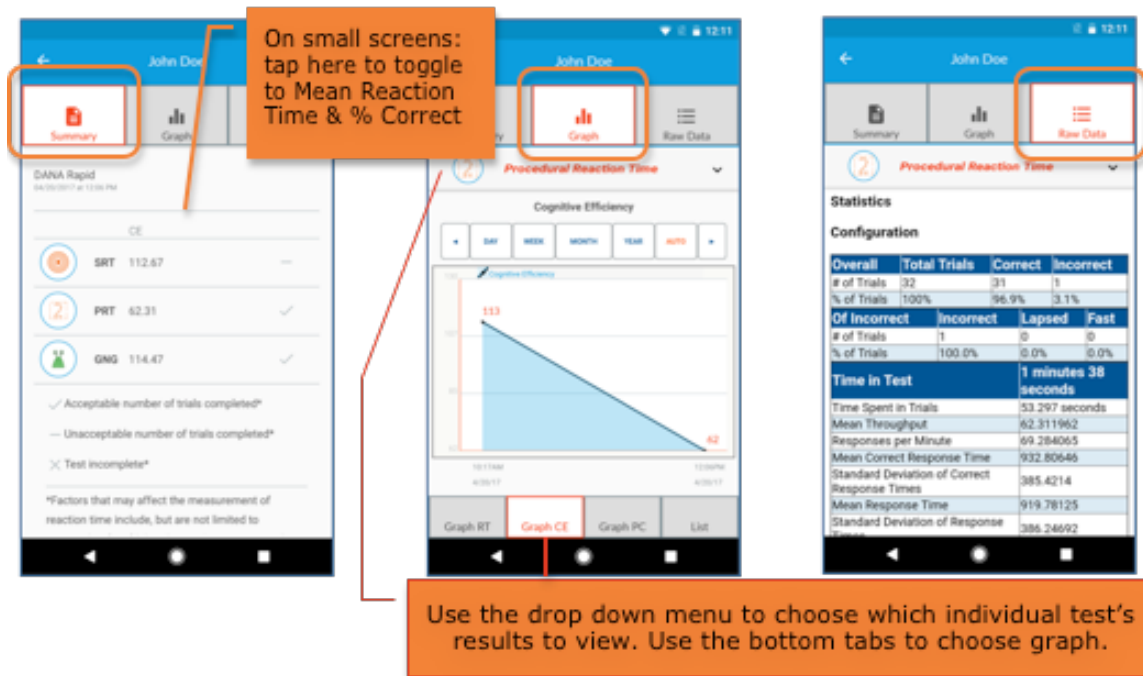 can adjust various test parameters, including number of practice and regular trials and the inter-trial interval. The default test batteries and tests are never deleted from the app via any customization; any custom test batteries or tests are simply added as additional screening options.

Objective 2: DANA DDM Opt-Out. Status: Completed

Task 1: Integrate cloud in DANA code base.

To improve usability, we added a cloud-based component to the DANA system. We implemented a relational database that, if hosted on an internet-enabled server, will automatically interface with the DANA mobile app and Web Portal to handle authentication (app and portal) and data uploads from the app to the database. Additionally, we created the DANA Web Portal front-end (mentioned under Task 3), which when also hosted on an internet-enabled server, provides an intuitive user interface to the data in the cloud database.

Task 2: Proper encryption for wireless transmission.

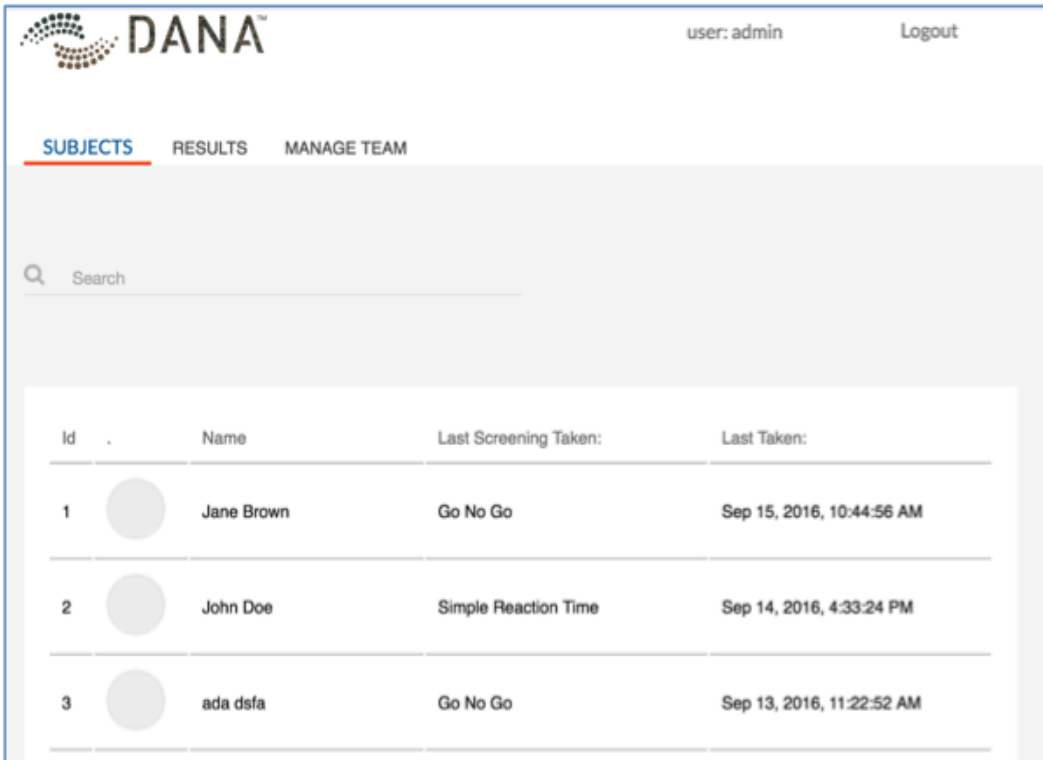We implemented transport layer security (TLS) encryption for security during any wireless data transmissions between the DANA app and the cloud database. Examples of such transmissions include (1) assessment data being uploaded from the DANA app to the cloud database and (2) communications in the opposite direction during cloud authentication (when logging in to the

Final Report (W81XWH-15-C-0141)

app). TLS is an industry standard security measure that implements cryptographic protocols to provide privacy and integrity of data between two computer applications.

Task 3: Develop dashboard for server.
We developed the DANA Web Portal dashboard as the front-end for the cloud database. The Web Portal has three main sections: (1) Subjects, (2) Results, and (3) Manage Team.
- o The **Subjects** section lists all subjects who have completed a DANA screening and had their data uploaded to the cloud database. Alongside each Subject, basic information about the last completed screening is listed – screening type, date and time of completion.
- o Once a subject is selected, the **Results** section is viewable. This is where all completed screenings for that subject are listed. Selecting a screening displays the summary report for that screening. From this section, you can also download data in either CSV (Extract CSV) or PDF (Download PDF) format.
- o The **Manage Team** section is where admin-level users can view and manage members of the administrative team: admins, clinicians, and examiners. This is where Admins can create additional users. See the DANA User Guide for the different permission levels for each user.

Task 4: Develop plug-ins for analysis tools. <u>Status: Completed</u>
We implemented data exports in CSV so that DANA data can be easily manipulated or imported into other analysis programs such as Microsoft Excel, SPSS, or R.

Left-to-right: Global export function, and export folder on device.

Final Report (W81XWH-15-C-0141)

CSV Export Results Imported into a Spreadsheet:

dana_data_1_SRT_2017_5_1_144202.csv

Test ID	Test Type GL	Subject GUID	Subject First	Subject Last	Battery GUID	Date	Time	Trial Number	Response	Response (N	Response Tir	Intertrial Int	Test Round	Assessment	DANA Versio	ITI Lower Bo	ITI Upper Bo	# of Practice
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	1	CORRECT	0	324	682	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	2	CORRECT	0	298	1757	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	3	CORRECT	0	287	2622	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	4	CORRECT	0	271	1245	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	5	CORRECT	0	294	1279	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	6	CORRECT	0	287	2944	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	7	CORRECT	0	295	2896	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	8	CORRECT	0	263	1635	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	9	CORRECT	0	319	1695	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	10	CORRECT	0	303	1248	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	11	CORRECT	0	344	972	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	12	CORRECT	0	329	1829	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	13	CORRECT	0	291	863	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	14	CORRECT	0	428	2800	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	15	CORRECT	0	281	2948	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	16	CORRECT	0	267	764	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	17	CORRECT	0	334	2371	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	18	CORRECT	0	277	1854	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	19	CORRECT	0	292	1365	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	20	CORRECT	0	328	2921	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	21	CORRECT	0	276	2137	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	22	CORRECT	0	763	1255	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	23	CORRECT	0	406	1809	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	24	CORRECT	0	303	2452	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	25	CORRECT	0	352	2708	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	26	CORRECT	0	334	1355	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	27	CORRECT	0	325	1493	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	28	CORRECT	0	308	1037	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	29	CORRECT	0	313	1782	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	30	CORRECT	0	311	2122	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	31	CORRECT	0	304	872	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	32	CORRECT	0	409	1665	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	33	CORRECT	0	373	1097	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	34	CORRECT	0	315	1937	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	35	CORRECT	0	306	1410	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5
1	A6A35ADBE	6140FB4608	John	Doe	E0DCE56B7E	5/1/17	14:36:22	36	CORRECT	0	331	620	0	DANA Rapid	4.0.0-RC.6-RI	600	3000	5

DATA ANALYSIS

Objective 3: Analyze and characterize existing datasets, and process-develop the CODE SAT algorithm, and

Objective 4: Validate the CODE SAT algorithm and implement it into the DANA codebase. Status: Both Completed

Task 1: Analysis of previously collected datasets.
To begin this task, we coded, cleaned, and formatted existing DANA datasets. Our initial dataset consisted of the following group-level differences: psychologically healthy vs. unhealthy ("Ft. Hood" data), hypoxic vs. non-hypoxic ("Altitude data"), concussed vs. non-concussed ("Air Force" data), and Alzheimer's Disease patients vs. normal elderly controls ("Burke" data). After applying candidate statistical techniques for the CODE-SAT algorithm on this combined dataset, we sought a new dataset that would yield a more robust division between normal and non-normal groups. Accordingly, we examined a dataset comprised of DANA data on concussed vs. non-concussed individuals. Around 210 college athletes were administered DANA to collect baseline data. Then, athletes were followed over the course of a season, and for those who sustained a concussion, DANA was re-administered at 24 hours, then at 8, 15 and 45 days post-injury. DANA was also administered at these time points to demographically matched controls. We applied the repeated measures, trial-by-trial techniques in an attempt to distinguish between concussed and non-concussed individuals. The results of this effort are summarized in the attached report. In the appendices: "Summary of trial-by-trial level analysis."

Final Report (W81XWH-15-C-0141)

- o Repeated measures manuscript was drafted: "*Cognitive Signatures: A Granular Approach to Studying Cognitive Efficiency*." (In appendices)

- o Ft. Hood manuscript "*Computerized cognitive testing norms in active-duty military personnel: Potential for contamination by psychologically unhealthy individuals*" was accepted for publication to Applied Neuropsychology: Adult (In appendices)

Task 2: Development of CODE-SAT algorithm.
The fundamental insight behind the CODE-SAT algorithm is the idea that longitudinal trial-by-trial response time profiles can be more sensitive to group differences than traditional summary measures (e.g., mean, standard deviation, etc.). To explore this possibility, we needed to determine (i) whether it was feasible that such differences in response profiles exist and (ii) how such differences could be mathematically characterized and formally tested. To examine feasibility, we fit loess curves to visually evaluate the shape and pattern of response profiles. After determining that profile differences could theoretically yield a basis for group differences, these loess curves were used as the targets for a parametric technique, linear spline regression. These models revealed significant differences between groups at particular time points. Then, to evaluate the effect of the overall shape of the response profile, we utilized a machine learning technique, *k*-means clustering, to categorize the data into separate groups on the basis of longitudinal profile shapes. With this analysis, new, out-of-sample subjects can be classified as belonging to one of these groups, where membership in certain groups can be indicative of non-normal cognition.

Task 3: Validate and implement CODE SAT.
The scope of the data analysis tasks was exploratory in that we were not certain if the approaches and/or data we had previously collected could serve as the best fit to validate the CODE SAT algorithm. At the beginning stages of the project, it was especially difficult to be certain that analyses of existing datasets could yield a fully validated algorithm. This is because the main theoretical insight, i.e., that differences in trial-by-trial response profiles could potentially signal cognitive impairment, was extremely novel. One particular challenge concerned whether the types of longitudinal patterns encountered in the data would be sufficient to generalize the concept of cognitive impairment to all other potential datasets.

With this challenge in mind, the CODE SAT algorithm could not be fully validated on the basis of our existing datasets. In particular, the limited number of binary classifications we considered (e.g., concussed vs. non-concussed, hypoxic vs. non-hypoxic, etc.) were unable to yield a general solution to the separation of "cognitively impaired" vs. "normal" individuals. While the CODE SAT algorithm could in theory be applied to specific separations such as concussion, the intent of this effort was to develop a general procedure for classifying impaired individuals, and our previous datasets were not sufficient to develop such an algorithm.

Despite these challenges we have included explicit instructions for the partially validated algorithm's implementation. Code written in the R programming language with comments to aid in running the algorithm will be provided.

Final Report (W81XWH-15-C-0141)

Future research efforts would focus on discovering which longitudinal patterns DANA data could be representative of cognitive impairment in general. Such work would require the collection of additional datasets comprising other sources of cognitive impairment (e.g., stroke, ADHD, etc.). These additional data would afford a more detailed comparison of individual impairments, which would hopefully yield a general trial-by-trial pattern indicating the presence of cognitive impairment.

TRANSITION

Objective 5: Transition DANA for Acquisition. Status: Completed

Task 1: Collaborate with stakeholders, provide briefings, and present and publish findings. Through the work performed under this Rapid Innovation Fund contract, AnthroTronix (ATinc) has made substantial progress in transitioning DANA to U.S. Special Operations Command (SOCOM). ATinc continued to engage with the U.S. Army MRMC's Non-Invasive Neuro-Cognitive Assessment Device (NINAD) Integrated Product Team (IPT). And, ATinc developed a detailed technology transition plan that will be very useful as it transitions DANA.

Since its initial meeting with SOCOM in August 2016, ATinc has worked closely with them to advance the transition of DANA. To perform an initial evaluation of DANA, SOCOM purchased an Android tablet and mobile phone with DANA pre-loaded on them. As a result of this evaluation, SOCOM developed plans to use DANA for a trial study, planned for Summer 2017, in conjunction with its selection classes. After its evaluation of DANA, SOCOM requested that ATinc make several changes to meet its operational needs, which ATinc completed under this contract. ATinc delivered this version of DANA to SOCOM (DANA 4.1.0-SOCOM) on April 12, 2017 SOCOM has asked ATinc to prepare a plan and proposal for more substantial changes to DANA, to include porting it to a Windows tablet and developing a seamless flow of data from DANA into the SPEAR database that SOCOM uses, which shows that SOCOM is thinking ahead to how DANA would be deployed downrange and integrated into its Concepts of Operations.

ATinc participated in the NINAD IPT Industry Day on December 7, 2016 in Baltimore, MD. ATinc answered initial follow up questions from the IPT by email and was notified by Mr. Brian Dacanay of the IPT that we will be contacted for an assessment of our commercialization strategy and manufacturing capabilities.

One of the deliverables under this contract is a technology transition plan for DANA. Having this plan will be extremely helpful to ATinc as DANA moves to transition. Through the process of writing this plan, ATinc synthesized information that it has gathered from its discussions with potential transition partners, such as SOCOM, and has identified key issues that will need to be addressed to ensure a successful transition of DANA. (In the appendices: "DANA Transition Package")

TRANSITION ACTIVITY BREAKDOWN BELOW:

Transition Activities from January 1, 2017 to March 31, 2017
DoD Researchers interested in DANA
- Sent quote for work related to DANA to LCDR Jay Haran, USN at the Submarine Medical Research Lab on February 1, 2017.
- Sent quote to Elizabeth Bergeron and John Florian at the Navy Experimental Diving Unit for work related to DANA on February 9, 2017.
- Scheduled meeting with Dr. Gary Kamimori at Walter Reed Army Institute for Research for April 6, 2017 to brief on DANA updates and to learn more about any potential funding.

Institute for Defense Analyses (IDA)
The Institute for Defense Analyses is interested in using DANA for a study they have proposed to a Pentagon sponsor, which would aim to evaluate the impact of using an unmanned ground robot, the Squad Mission Equipment Transport vehicle, on the cognitive and physical performance of an infantry squad. This study would occur at Ft. Benning; IDA is responding to their sponsor's request for additional information, and they hope to have a decision regarding funding by May 31, 2017. If they are funded, they would plan on purchasing between 10-25 tablets with DANA loaded on them.

Transition Plan
Completed the DANA Technology Transition Plan, which is included with this report. The transition plan is included in the appendices labeled "DANA Transition Package."

SOCOM
- On March 31, 2017, DANA 4.1.0-SOCOM was completed; this version includes additional features that SOCOM requested be included in DANA. SOCOM sent the Android tablet they purchased back to ATinc where it was loaded with the new DANA version. (The Android phone was in use by SOCOM at another facility at the time, so was not available at the time to send back for the update.)
- Multiple conference calls were held with Travis Harvey of the Protection of the Force and Family within SOCOM regarding SOCOM's interest in exporting the data from DANA into their human performance database, SPEAR. Based on feedback that ATinc received from Mr. Harvey, ATinc prepared a high-level proposal and cost estimate to (a) demonstrate DANA data integration into SPEAR after a Bluetooth data transfer from an Android tablet running DANA to a Windows 10 tablet running SPEAR, and (b) develop a Windows 10 version of DANA that could run on the same Windows tablet as SPEAR. Mr. Harvey believes that for DANA to fit into SOCOM's Concept of Operations (CONOPS), it needs to run on the same device as the SPEAR database. We believe that this demonstrates SOCOM's serious interest in acquiring DANA for deployment down range. We delivered this high-level proposal and cost estimate to Mr. Harvey for his review.

Final Report (W81XWH-15-C-0141)

Other
- Attended NDIA Military Medicine Partnership Conference on March 7-8, 2017.

Transition Activities from October 1, 2016 to December 31, 2016
Institute for Defense Analyses (IDA)
- Meeting with Jim Kurtz and his colleagues on November 9, 2016.
- They just completed one study looking at the effects of physical stress on dismounted soldiers and are very interested in DANA as means of assessing impact on cognitive processing.
- In their next study, they would plan on using DANA to compare cognitive processing of dismounted infantry squads using an unmanned Squad Mission Equipment Transport (SMET) class vehicle, which has a cargo capacity of approximately 1,500 lbs., vs. squads that are not using those vehicles to carry their loads.
- If IDA gets funded to continue their study, they are interested in purchasing 8-10 Tablets with DANA loaded on them, based on a quote we provided to MAJ Mike Dretsch of TRADOC, as noted below.
- IDA briefed their Pentagon sponsor on their study on December 16, 2016, and hoped to get an indication of future funding at that time.
- Follow up with them regarding their funding status from their Pentagon sponsor.

MRMC
- Participated in kick off meeting for DoD Healthy Brain proposal we were invited to submit.
- Reviewed draft Impact Statement written by Dr. Timothy Lacy, Senior Medical Advisor to AnthroTronix, and draft proposal for the "DoD Caregiver Study" pre-proposal that we submitted to the Peer Reviewed Alzheimer's Research Program. Submitted a proposal on November 8, 2016
- Participated in the Non-Invasive Neurocognitive Assessment Device (NINAD) Integrated Product Team (IPT) Industry Day meeting on December 7, 2016 at the Inner Harbor Baltimore Marriott, with Dr. Tim Lacy.

MUSTER
Received a SBIR Phase II Option contract of $750,000 from the Office of Naval Research for the PASS MUSTER project, which utilizes DANA in conjunction with physical vital signs/biometrics.

SOCOM
- Shipped COL Mark Baggett, Command Psychologist for SOCOM, an Android tablet and phone running DANA. SOCOM received the tablet and phone running DANA that were shipped.
- Met with COL Baggett and a number of his colleagues at SOCOM headquarters at MacDill AFB, FL on November 10, 2016.
- Planning on ordering 25 tablets so that he can conduct normative study with 2,000 test subjects once IRB is approved.
- Sent SOCOM all published papers on DANA.

Final Report (W81XWH-15-C-0141)

- Held a conference call with COL Baggett and Ed Deagle of SOCOM on December 12, 2016.

TRADOC
- Sent quote to MAJ Mike Dretsch, Chief Cognitive Scientist for US Army TRADOC on Oct. 21st for 20 tablets and phones running DANA, which would be used in conjunction with the study IDA might perform, as noted above.

Other
Completed new product sheet for DANA Military, which shows updated graphics, draft summary test reports and DANA configurations.

Transition Activities from July 1, 2016 to 30 September 30, 2016
- At the Military Health Systems Research Symposium, held in Orlando, FL on 15-17 August, met with**:**
    o SOCOM: COL Baggett Psychologist, CAPT Cota, Command Surgeon, and LTC Nuce
    o MRMC: Christie Vu, MAJ Carr, and Brian Dacanay
    o DHA: CMDR Joseph Cohn
    o TATRC: Jim Beach, COL (Ret) USA
    o Various researchers, including Gary Kamimori and Lt. Jay Haran, USN
- Submitted response to the RFI for Non-Invasive Neurological Assessment Devices (NINAD) for detecting mild-to-moderate mTBI on September 1, 2016 one day ahead of schedule.
- Submitted response to the RFI for Non-Self Reporting Methods for Detecting Changes in Psychological Status from JPC5 on September 12, 2016.
- Received trial order from COL Baggett at SOCOM for one Android-based tablet and phone running DANA; shipped the order to COL Baggett in October.
- Scheduled meeting at MacDill AFB with COL Baggett, CAPT Cota, ad LTC Nuce for November 10, 2016.
- Met with Regina Shia, researcher from AFRL 711th Human Performance Wing on September 7, 2016.
- Conference call with Jim Kurtz from Institute for Defense Analysis on September 23, 2016; TS Jones, MG (Ret) USMC referred them to us. IDA is nearing the end of a study looking at the effects of physical stress on dismounted Infantry, and from my initial conversation with Jim are interested in DANA. Scheduled follow up meeting with Jim for November 9, 2016.
- Attended AUSA's "Hot Topics" meeting on Army medicine on 22 September

Transition Activities from April 1, 2016 to June 30, 2016
- Held conference call with CDR Joseph Cohn, Director, Advanced Development Program Research, Development, and Acquisition Directorate, Defense Health Agency on April 5, 2016 regarding transitioning DANA in conjunction with JPC-5. Prepared brief for CDR Cohn to share with JPC-5 and other colleagues.
- Jonathan Brown attended National Defense Industrial Association (NDIA) on "Medical Research, Development and Acquisition" on 18-20 April in Ellicott City, MD.

Final Report (W81XWH-15-C-0141)

- Met with Baruch Ben Dor, CEO of InfraScan, on June 9, 2016 to discuss concept of integrating DANA with the InfraScanner 2000, which is a hand-held device using Infrared technology to detect brain hematomas
- Included DANA as one of the metrics ATinc proposes to use in our role supporting Human Experimentation as part of Raytheon BBN's proposal to DARPA for the Squad X program. Proposal submitted to Raytheon BBN on June 23, 2016. Raytheon BBN will be submitting their proposal to DARPA on July 15, 2016. However, DARPA did not select the Raytheon BBN proposal.
- On June 27, 2016 ATinc submitted pre-application titled "Creating a mobile app to build cognitive resilience to stress-related impairment" in response to the "Cognitive Resilience and Readiness Research Program Announcement. Team members for this proposed effort include leading researchers on stress and sleep deprivation from the University of Pennsylvania and University of California, San Diego.
- Conference call on June 30, 2016 with TS Jones, MajGen, USMC (Ret'd.) regarding study he is currently conducting looking the cognitive resilience of warfighters for the Office of the Secretary of Defense. Scheduled meeting with TS Jones for July 11, 2016.
- Abstract titled "Trial-by-Trial Pattern Analysis: A Novel Strategy for Identifying Neurocognitive Deficit with Computerized Cognitive Tests" was accepted as a poster presentation at the Military Health Systems Research Symposium (MHSRS).

Transition Activities from January 1, 2016 to March 31, 2016
- Tim Lacy attended "Brain Health Summit," on January 20, 2016 held by COL Benjamin Solomon, MD, Brain Health Program Manager, System for Health Directorate and Performance Triad, Deputy Chief of Staff for Public Health, Office of The Surgeon General.
- Mr. Jonathan Brown, Business Development Consultant for AnthroTronix, attended National Defense Industrial Association (NDIA) Human Systems Conference on February 9-10, 2016 in Springfield, VA.
- Dr. Corinna Lathan, Ms. Charlotte Safos, Chief Operating Officer at AnthroTronix, Jonathan Brown, and Tim Lacy met with The Phoenix Group on February 10-11, 2016. The Phoenix Group is one of ATinc's partners for entering the U.S. commercial market with DANA.
- Held conference call with CDR Reese of the U.S. Navy Bureau of Medicine and Surgery on February 26, 2017.
- Jonathan Brown and Tm Lacy met with COL Sid Hinds, DVBIC's outgoing Director, and his Directors of Education, Research, and Clinical on February 22, 2016.
- Jonathan Brown gave a presentation on using DANA to assess changes in Cognitive Processing at Global Force Symposium, sponsored by the Association of the United States Army (AUSA) in Huntsville, AL on March 15, 2016. LTG Kevin Mangum, the Deputy CG of TRADOC, attended the talk and seemed engaged.
- ATinc submitted proposal titled "Completing the Transition of the Defense Automated Neurobehavioral Assessment (DANA) to Operational Use" to Joint Warfighting Medical Research on March 30, 2016. Collaborators for this project include researchers Johns Hopkins Medical School.

Final Report (W81XWH-15-C-0141)

Transition Activities from October 1, 2015 to December 31, 2015
- Met with Pet Palmer, Director, General Dynamics Edge Network, at AUSA regarding Cognitive Readiness effort.
- Call with Cori and Dr. Christie Vu on October 21, 2015 regarding the Joint Warfighter Medical Research Program.
- Met with Brian Dacanay, co-chair of the NINAD IPT on October 22, 2015.
- Conference call with LTC Scott Moran, ImPACT Program Manager USASOC on October 29, 2015.
- Attended the Neurological Behavioral Health Subcommittee meeting of the Defense Health Board on "Scientific Evidence of Using Population Normative Value for Post-Concussive Computerized Neurocognitive Assessments" at MacDill AFB in Tampa, FL on November 9, 2015.
- Conference call on November 24, 2015 with MAJ Carr, MAJ Yarnell, and Thomas Baker re questions to get CRADA in place so that WRAIR can share data with us
- Attended meeting hosted by COL Ben Solomon, Brain Health Program Manager, in the Office of the Army Surgeon General, at DHA on December 2 and 3, 2015.

What opportunities for training and professional development has the project provided?

Over the course of two years, our staff has had the opportunity to learn about and utilize their knowledge in the following technologies:

- NodeJS and AngularJS technologies, which are used to implement the DANA cloud database and DANA web portal
- Android and Android Studio
- ProGuard (Android code obfuscation tool)
- Photoshop
- JIRA and Bitbucket

Additionally, our staff attended Android development conferences and developer meet-ups.

Over the course of developing the CODE SAT algorithm, there were multiple opportunities for the training and professional development of those involved. The CODE SAT algorithm required the utilization of advanced statistical methodologies, and the team members working on this aspect had the opportunity to learn to apply these methods. This involved both understanding the techniques from a theoretical and mathematical perspective as well as learning how they are implemented on computerized platforms. In addition to understanding individual techniques, team members needed to learn how to evaluate each one relative to the others, discovering the strengths and weaknesses of each. More generally, a significant professional development opportunity was presented by the challenge of taking a highly theoretical idea about cognitive performance and turning it into a tangible reality, a broad exercise that a single team rarely sees completed from start to finish.

Final Report (W81XWH-15-C-0141)

How were the results disseminated to communities of interest?
We have compiled a citation list of all relevant DANA publications, which are now available online at danabrainvital.com/research with a copy of the publication. During the reporting period, we also published an additional DANA related paper and poster. (Please see products section and appendices)

1. Hollinger, K. R., Franke, C., Arenivas, A., Woods, S. R., Mealy, M. A., Levy, M., & Kaplin, A. I. (2016). Cognition, mood, and purpose in life in neuromyelitis optica spectrum disorder. Journal of the neurological sciences, 362, 85-90.
2. Lathan CE, Coffman I, Shewbridge R, Lee M, Cirio R, et al. A Pilot to Investigate the Feasibility of Mobile Cognitive Assessment of elderly patients and caregivers in the home. J Geriatrics Palliative Care 2016;4(1): 6.
3. Resnick, H., & Lathan, C. (2016). From battlefield to home: a mobile platform for assessing brain health. MHealth, 2(7). Retrieved from http://mhealth.amegroups.com/article/view/11037
4. Steeg Morris, D., Lathan, C., Weissfeld, L, Lacy, T., & Resnick, H. R. (2016, August). Trial-bytrial pattern analysis: A novel strategy for identifying neurocognitive deficit with computerized cognitive tests. Poster presented at the Military Health System Research Symposium, Orlando, FL.

What do you plan to do during the next reporting period to accomplish the goals?
- Nothing to report; this is the final report.

Final Report (W81XWH-15-C-0141)

IMPACT

What was the impact on the development of the principal discipline(s) of the project?
The utility of neurocognitive assessment tools (NCATs) crucially depends on what can be learned from the data they collect on cognitive performance. Most currently available NCATs take a relatively simple approach by collecting information only on summary performance measures, such as one's average reaction time on a certain cognitive test. We are enhancing the utility of NCATs by looking at alternative, potentially more informative ways of interpreting the data they collect. The work we have done in this area involves understanding how different ways of looking at the data, in conjunction with the appropriate mathematical techniques to analyze them, can tell us more about cognitive impairment over the techniques implemented in most NCATs. Thus, the primary impact our work has made on the field of NCATs is the ability of these devices to both detect and monitor changes in cognitive performance with a potentially much greater degree of precision. In addition to improving the utility of DANA, we hope to establish a new precedent in the interpretation of neurocognitive data that will be appreciated by the field as a whole.

Moving forward, we will focus less on the CODE SAT algorithm's ability to distinguish between groups (i.e. impaired vs. unimpaired) and instead direct our efforts towards using the algorithm to detect within-patient, longitudinal changes. While most extant NCATs can detect meaningful cognitive differences at a single time point, none have been developed with the explicit goal of applying advanced analytical techniques to longitudinal data. The impact of this work on the discipline of neurocognitive assessment would be substantial; NCATs would be able to detect extremely subtle yet meaningful changes in cognitive performance for a variety of applications, including treatment response (e.g., after stroke or concussion) and intra-operative monitoring, for example. In addition to such clinical applications, the CODE SAT algorithm would allow academic researchers in the cognitive sciences to potentially uncover new knowledge about cognitive functioning when the data are able to represent extremely subtle changes of theoretical interest.

What was the impact on other disciplines?
Our published work under this contract impacts the practice of those performing research on and/or utilizing neurocognitive assessment tools (NCATs) in practical settings. We highlight two recent publications in support of this assertion: a presentation on our "trial-by-trial level" analysis work presented at the Military Health System Research Symposium (MHSRS) and a broader perspectives piece published in the journal mHealth. Our MHSRS presentation summarized our findings related to work on trial-by-trial-level analyses of neurocognitive data, showing the potential of this mode of analysis to yield richer, more informative insights over traditional analyses that rely only on aggregate, summary measures of response time data. Our mHealth publication, titled "From battlefield to home: A mobile platform for assessing brain health," takes a broad perspective on the role of computerized cognitive testing and highlights the unique ability of NCATs' computerized (rather than pencil-and-paper) platform as well as a potential shift away from infrequent health measures to rich, longitudinal data facilitated by the ease of NCATs implementation in mobile devices.

Many fields are potentially impacted by these findings. For example, both of the

Final Report (W81XWH-15-C-0141)

abovementioned publications have clear utility in the field of neurocognitive/neuropsychological testing. The insights yielded can both (a) allow clinicians to better understand the data resulting from computerized neurocognitive assessments (by way of trial-by-trial level analysis), and (b) have a more accurate understanding of patients' cognitive health (via high-frequency testing at home on a mobile device). In addition to the clinical setting, higher-level research in neuropsychology may benefit as well. In particular, trial-by-trial-level analyses lend themselves to new statistical/mathematical methods that can help researchers understand the nature of particular neurocognitive impairments rather than just signal their existence. Because our work on NCATs has been of a fundamental nature, the potential to impact fields beyond those concerned with neurocognitive testing in general is great.

What was the impact on technology transfer?
As noted in our transition activity and the DANA Transition Package, AnthroTronix has made substantial progress with US Special Operations Command (SOCOM), with respect to transition. The DANA Transition Package details the specific plan to transition DANA with SOCOM based on what we have learned from our meetings and conference calls with them. We better understand SOCOM's specific needs with respect to product configurations, CONOPS, as well as how they would need DANA integrated into their health information systems.

What was the impact on society beyond science and technology?
The results of our efforts to develop the CODE SAT will likely have an impact beyond the bounds of the science, engineering, and academic world. The primary goal of the CODE SAT algorithm is to detect, with unprecedented precision, cognitive impairment - an outcome that will be appreciated by diverse applications. For example, we focused on the detection of concussion. The CODE SAT algorithm can thus be utilized in formal clinical settings, sports medicine practitioners, etc. More generally, the mobile platform-based nature of DANA means that data-analytic solutions can be applied to any setting where a tablet/smartphone and internet connection are available, meaning that even ordinary individuals can use the DANA software along with the CODE SAT algorithm for a better understanding of their own cognitive health. In addition, many policy decisions relate to neurocognitive outcomes, ranging from pre-deployment testing for service members to policies on the driving ability of individuals suffering cognitive decline. With the CODE SAT algorithm, these issues can be considered in the context of a device that affords greater precision in assessing cognitive health outcomes.

Final Report (W81XWH-15-C-0141)

CHANGES / PROBLEMS
- Nothing to report

PRODUCTS
Publications in appendices
1. Hollinger, K. R., Franke, C., Arenivas, A., Woods, S. R., Mealy, M. A., Levy, M., & Kaplin, A. I. (2016). Cognition, mood, and purpose in life in neuromyelitis optica spectrum disorder. Journal of the neurological sciences, 362, 85-90.
2. Lathan CE, Coffman I, Shewbridge R, Lee M, Cirio R, et al. A Pilot to Investigate the Feasibility of Mobile Cognitive Assessment of elderly patients and caregivers in the home. J Geriatrics Palliative Care 2016;4(1): 6.
3. Resnick, H., & Lathan, C. (2016). From battlefield to home: a mobile platform for assessing brain health. MHealth, 2(7). Retrieved from http://mhealth.amegroups.com/article/view/11037
4. Steeg Morris, D., Lathan, C., Weissfeld, L, Lacy, T., & Resnick, H. R. (2016, August). Trial-bytrial pattern analysis: A novel strategy for identifying neurocognitive deficit with computerized cognitive tests. Poster presented at the Military Health System Research Symposium, Orlando, FL.

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

1. Name: Dr. Corinna Lathan
   Project Role: Principal Investigator
   Nearest person month worked: 1
   Contribution to Project: Dr. Lathan has provided direction for this effort and coordinated internal and external input.

2. Name: Clifford Knoll
   Project Role: Software Engineer
   Nearest person month worked: 8
   Contribution to Project: Mr. Knoll wrote new and modified existing DANA code.

3. Name: Ian Coffman
   Project Role: Research Scientist
   Nearest person month worked: 8
   Contribution to Project: Mr. Coffman conducted and managed on-going DANA data analysis and developed various data reports.

4. Name: Marissa Lee
   Project Role: Research Coordinator
   Nearest person month worked: 6
   Contribution to Project: Ms. Lee supported project management activities and coordinated DANA data management.

5. Name: James Drane
   Project Role: Technical Lead

Final Report (W81XWH-15-C-0141)

Nearest person month worked: 5
Contribution to Project: Mr. Drane led all technical development efforts and device and software quality assurance testing.

6. Name: Rita Shewbridge
   Project Role: Project Manager
   Nearest person month worked: 5
   Contribution to Project: Ms. Shewbridge coordinated the technical and research arms of this effort and performed all management activities.

7. Name: Sarah Staines
   Project Role: Research Assistant
   Nearest person month worked: 1
   Contribution to Project: Ms. Staines assisted with software documentation and quality assurance testing.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
   • Nothing to report

What other organizations were involved as partners?
   • Resnick, Chodorow and Associates were involved as statistical analysis partners.

Describe the Regulatory Protocol and Activity Status (if applicable).
(a) Human Use Regulatory Protocols

TOTAL PROTOCOLS: No human subjects research was performed for the above tasks; however, the purpose of this research was to analyze existing DANA data to discover unique characteristics of psychological impairment and cognitive deficiency.

Because we were analyzing existing data we were asked by ORP to provide IRB documentation for all of the data being used. The AnthroTronix IRB determined that the protocols were exempt and ORP HRPO agreed with this decision. *(Please see statement below)*

"The AnthroTronix Institutional Review Board (IRB) Office determined that the protocol is exempt as it is research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified directly or through identifiers linked to the subjects.

As required by DOD Instruction 3216.02, encl 3, paragraph 4.c(1), the ORP HRPO concurs with the exempt determination made by the AnthroTronix IRB Office. The project may proceed with no further requirement for review by the HRPO. The HRPO protocol file will be closed."

Final Report (W81XWH-15-C-0141)

DELIVERABLES
All deliverables will be submitted on Tuesday May 16, 2017 via FedEx on a CD, tracking number is 7791 5006 0944. The items included on the disk are as follows:

- Android app installable file (DANA 4.0.0-RIF)
- Web server repository + instructions
- Web portal repository + instructions
- DANA User Guide
- DANA Transition Package
- DANA CODESAT
- Final Report

SPECIAL REPORTING REQUIREMENTS
- Quad Chart

REFERENCES

Lathan C, Spira JL, Bleiberg J, Vice J, Tsao JW. Defense Automated Neurobehavioral Assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool. *Mil. Med.* 2013; 178 (4): 365–71.

Spira, James L., et al. "The Impact of Multiple Concussions on Emotional Distress, Post-Concussive Symptoms, and Neurocognitive Functioning in Active Duty United States Marines Independent of Combat Exposure or Emotional Distress." *Journal of neurotrauma* 31.22 (2014): 1823-1834.

Subudhi, Andrew W., et al. "AltitudeOmics: the integrative physiology of human acclimatization to hypobaric hypoxia and its retention upon reascent." PloS one9.3 (2014): e92191.

Roach, Emma B., et al. "AltitudeOmics: Decreased reaction time after high altitude cognitive testing is a sensitive metric of hypoxic impairment."NeuroReport 25.11 (2014): 814.

Russo, C. R., and C. E. Lathan. "An Evaluation of the Consistency and Reliability of the Defense Automated Neurocognitive Assessment Tool." *Applied Psychological Measurement* (2015): 0146621615577361

APPENDICES

- Quad Chart

- Data Analysis Reports
    - Simple Reaction Time Repeated Measures Analysis: Phase I Report
    - Simple Reaction Time Repeated Measures Analysis: Phase II Report
    - Repeated Measures Analysis: Phase III Report

Final Report (W81XWH-15-C-0141)

- o Repeated Measures Analysis: Phase IV Report: Analysis of four datasets Ft. Hood, Air Force, Altitude, and Burke (Healthy vs. Unhealthy)
  - o Repeated Measures Analysis: Phase V Report
  - o Summary of trial-by-trial level analysis

- DANA Manuscripts
  - o Repeated measures manuscript "*Trial-by-Trial Analysis: A New Approach to Detecting Neurocognitive Changes With Computerized Cognitive Tests.*"
  - o Ft. Hood manuscript "*Computerized cognitive testing norms in active-duty military personnel: Potential for contamination by psychologically unhealthy individuals*"

- DANA User Guide

- DANA Transition Package

- DANA CODESAT (On CD)

- DANA Publications
  1 Hollinger, K. R., Franke, C., Arenivas, A., Woods, S. R., Mealy, M. A., Levy, M., & Kaplin, A. I. (2016). Cognition, mood, and purpose in life in neuromyelitis optica spectrum disorder. Journal of the neurological sciences, 362, 85-90.
  2 Lathan CE, Coffman I, Shewbridge R, Lee M, Cirio R, et al. A Pilot to Investigate the Feasibility of Mobile Cognitive Assessment of elderly patients and caregivers in the home. J Geriatrics Palliative Care 2016;4 (1): 6.
  3 Resnick, H., & Lathan, C. (2016). From battlefield to home: a mobile platform for assessing brain health. MHealth, 2(7). Retrieved from http://mhealth.amegroups.com/article/view/11037
  4 Steeg Morris, D., Lathan, C., Weissfeld, L, Lacy, T., & Resnick, H. R. (2016, August). Trial-by-trial pattern analysis: A novel strategy for identifying neurocognitive deficit with computerized cognitive tests. Poster presented at the Military Health System Research Symposium, Orlando, FL.

An Assessment Tool to Detect Unique Characteristics of Cognitive Deficiency

Log Number: RIF14037
Award Number: W81XWH-15-C-0141

PI: Dr. Corinna Lathan   **Org:** AnthroTronix

Award Amount: $995,000

Study/Product Aim(s)

- Analyze existing data to find unique impairment characteristics
- Update DANA technical features
- Transition DANA to military customers and programs

Approach

Develop an algorithm that can identify normal parameters in cognitive functioning using pre-existing DANA cognition data. Create a drag and drop user interface (UI) for DANA and update subtest code to integrate this feature and eliminate the need for the DDM by creating a cloud-based system. Work with stakeholders to position DANA for operational use.
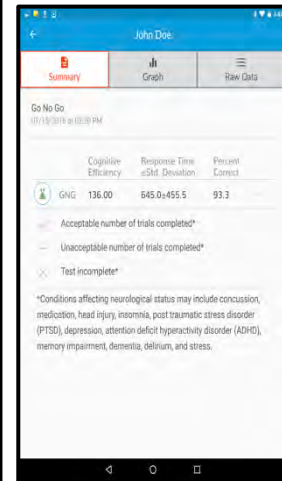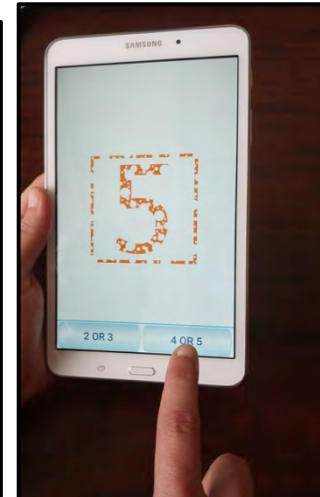
Accomplishments: Cloud based system for storing DANA data has been completed; Batteries are fully customizable; Analysis for CODE-SAT development has been completed.

Timeline and Cost

Activities CY	2015	2016	2017
Customizable Batteries		███	
DANA DDM Elimination		███	
Transition DANA			███
Develop, validate, and implement CODE-SAT algorithm			███
Estimated Budget ($K)	**$198,000**	**$725,032**	**$71,968**

Updated: (May 15, 2017)

Goals/Milestones

CY15 Goal – Initial technical development of batteries and cloud
- ☑ First demo of drag and drop UI for customizable batteries
- ☑ Cloud database – a DANA result can be added to database

CY16 Goals – Final technical development of batteries; full transition to cloud-based data management; analyze existing DANA data
- ☑ Create desktop UI, dashboard, and analysis tool plugins for cloud
- ☑ Finalize drag and drop UI by end of January
- ☑ Test all new DANA technical features and eliminate any bugs
- ☑ Analyze cognitive performance DANA data and create algorithm from any characteristic patterns identified

CY17 Goal – Transition DANA and finalize CODE-SAT
- ☑ Validate algorithm and embed in DANA codebase
- ☑ Create updated user manual and full transition package

Comments/Challenges/Issues/Concerns
- None

Budget Expenditure to Date
Projected Expenditure: $995,000
Actual Expenditure: $995,000

Simple Reaction Time Repeated Measures Analysis: Phase I Report

Contents

List of Figures

List of Tables

1 Statement of the Challenge

Although much has been published on summary statistics of multi-trial cognitive function tests, little has been done to leverage all the trial data upon which these summary measures are based. It may be possible to use these data in ways that can improve on current strategies to identify people with head injury, depression, PTSD, and age-related cognitive decline. A first step in this process is to identify robust methods that describe patterns in trial data such that "normal" individuals can be distinguished from those outside the normal range.

2 Data

This report presents repeated measures analysis of the "SRT - Altitude" data from the Excel file "Altitude and Air Force Trial by Trial Data(updated).xlsx" received by Resnick, Chodorow and Associates on April 21. This first phase of the analysis makes some simplifying data assumptions:

- Only trials from administration 1 are included in the analysis. Given that the Air Force data is available only for administration 1, we made this cut to the Altitude data in order for the analysis to be comparable for Phase 2 analysis of the Air Force data.

- Lapsed trials ("response = Lapse') are excluded from the analysis.

- Only trials at 5260m above sea level ("altitude = 1") and at sea level ("altitude = 3") are included in the analysis.

After these exclusions we are left with a dataset of $1,462$ observations uniquely identified by ID, trial number, and altitude. We focus our attention on these three variables in addition to reaction time, the primary variable of interest. Table 1 presents summary information about the analysis data set.

Table 1: Summary Information about the Altitude Data

	Above Sea Level	Sea Level
Number of Subjects	21	17
Average Number of Trials Per Person	37.4	39.8
Range of Number of Trials Per Person	10-40	38-40
Average Response Time	337.1	299.0
Range of Response Time	160-863	198-757
Average of Subject-Average Response Time	344.9	299.3
Range of Subject-Average Response Time	261.9-543.2	266.9-380.1

3 Exploratory Data Analysis

Exploratory data analysis of the trial profiles of response time is a crucial step in understanding the mean and variance structure of the data. The analyses presented in this report focus on uncovering trends and characteristics in the evolution of mean response time over the course of administering

the multiple trials of the simple reaction time test. The following exploratory graphs and statistical methods were used to select an appropriate mean structure for a linear model of the trial profiles.

Figure 1 plots the trial profiles of response time for each subject by altitude. Each line represents one subject's response times traced over up to 40 trials. The trial profiles appear erratic in that within altitude, there is no obvious visual trend that is common across subjects. While individual profiles are extremely difficult to distinguish in Figure 1, some average features are more apparent across the two altitudes: (1) above sea level response time measurements appear more variable, and (2) the bulk of the above sea level measurements seem to lie above the bulk of the sea level measurements.
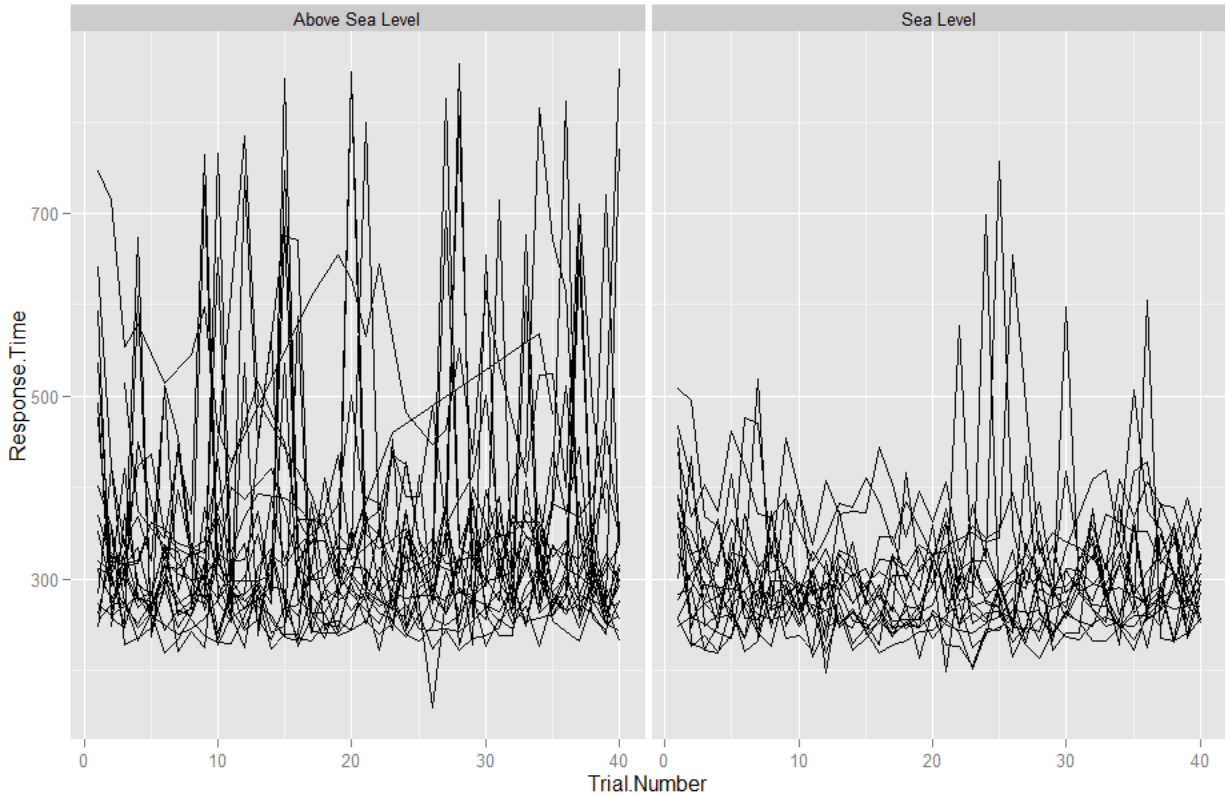

Figure 1: Trial Profiles of Response Time by Altitude

Figure 2 adds average response times by trial number (red triangles) to the trial profiles shown in Figure 1. The triangles confirm that the sea level trial averages generally lie below the above sea level averages. Additionally, the sea level trial averages form a relatively stable trend line. This contrasts with the more erratic averages that are observed in the above sea level subjects.
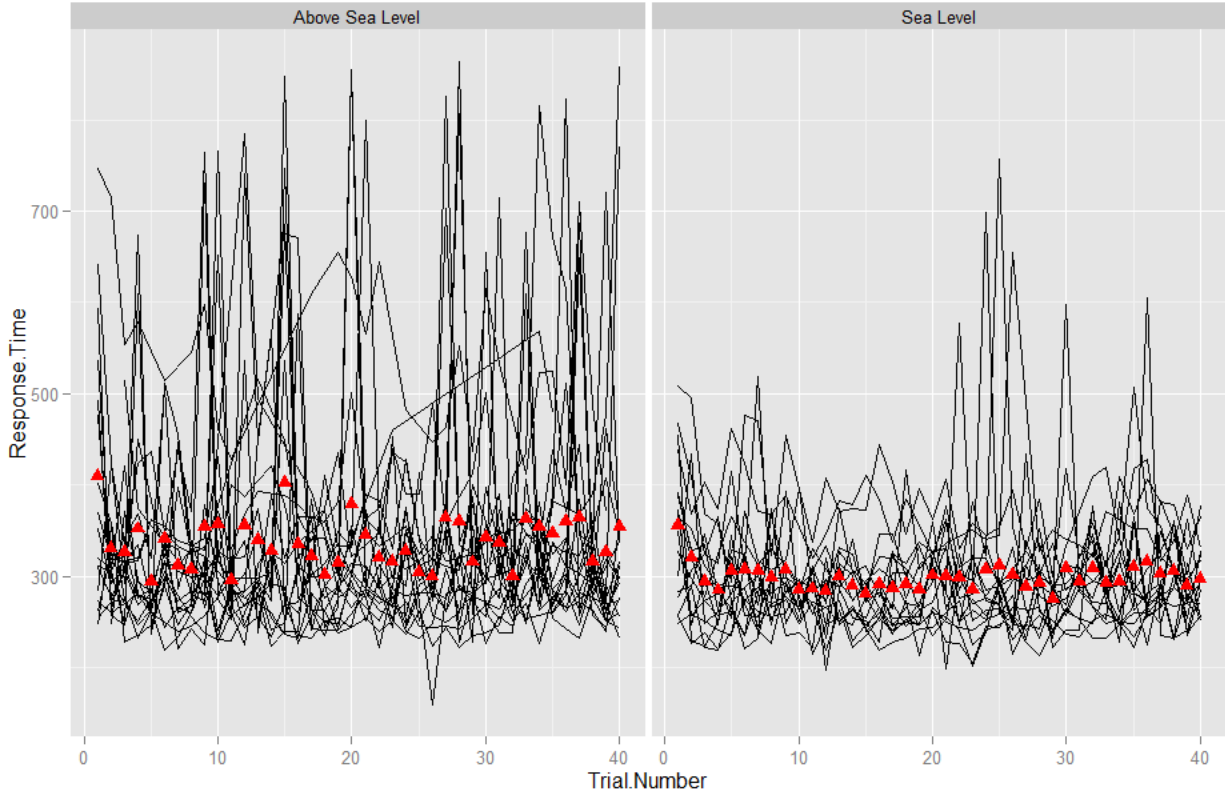
Figure 2: Trial Profiles by Altitude: Trial Averages

Figure 3 includes a loess curve fit to the trial data. A loess curve is a type of non-parametric smooth fit that is data-driven and empirically derived. The loess does not require any a priori model specification, nor does it rely on parametric assumptions about the shape of the trend. This technique estimates a curve by fitting multiple simple models to localized ranges of the x-axis, and it provides an appealing graphical summary of the relationship between response time and trial number. Imposing loess curves on the plots of trial profiles facilitates visualization of differences in their means and variability: the loess curve for sea level subjects hovers around 300, while the loess curve for the above sea level subjects exhibits wigglier behavior that is frequently larger than 300. In addition, average response times by trial (the red triangles) are more variable around the loess curve for the above sea level subjects than for the sea level subjects.

The loess curves also provide insight into appropriate models for average response times across trials. Despite the apparent variability in the above sea level subjects measures, the loess curves are remarkably linear with the most pronounced curvature in the first 0-15 trials or so. After about trial 5 the sea level subjects' average profiles are strikingly linear, while curvature remains evident among the above sea level subjects' average profiles.
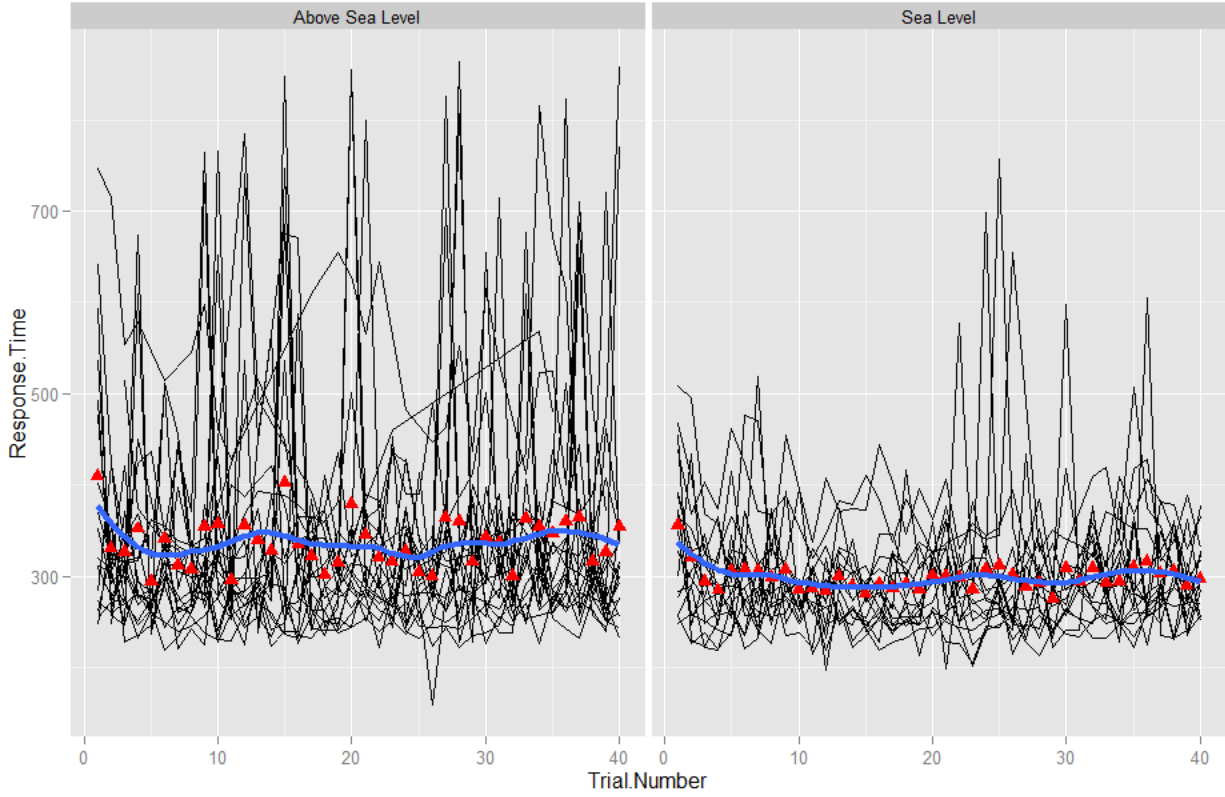
4

Figure 3: Trial Profiles by Altitude: Loess

4 Linear Models

Using observations from the exploratory data analysis and insight concerning the shape of the curve from the loess, we fit linear statistical models to capture the dependence between response time and trial number. Models of the two altitudes were assessed separately with the goal of identifying a generally applicable model.

4.1 Quadratic Regression

As noted previously, both trends look roughly linear with curvature in some ranges of trial number. A simple model for such a trend is quadratic regression: a simple linear regression that includes a quadratic term for trial number. We observed that this model captures the appropriate amount of curvature in sea level subjects, but not in the above sea level subjects. Table 2 presents coefficient estimates, standard errors and statistical significance for the quadratic regression strategy. The table shows that trial number and its quadratic term are only statistically significant in modeling the sea level subjects' profiles. The lack of statistical significance for the above sea level subjects' profiles is likely due to model misspecification. That is, the quadratic terms are not adequately capturing the curvature. Figure 4 shows that quadratic polynomial linear regression over-smooths

5

(particularly for the first 15 trials) and masks the loess curvature in the above sea level case while it provides a relatively good fit for sea level subjects. Considering all the evidence, quadratic regression does not appear to be a modeling tool that applies well to subjects at both altitude levels.

Table 2: Quadratic Regression Results by Altitude

Covariate	Model 0 Estimate (s.e.)	Model 1 Estimate (s.e.)	Model 2 Estimate (s.e.)
	Above Sea Level		
Intercept	337.1 (4.10)**	335.8 (8.39)**	347.3 (12.96)**
Trial Number		.06 (.36)	-1.59 (1.46)
Trial Number Squared			.04 (.03)
	Sea Level		
Intercept	299.0 (2.37)**	301.9 (4.83)**	317.1 (7.44)**
Trial Number		-.14 (.21)	-2.32 (.84)**
Trial Number Squared			.05 (.02)**

Note: Statistical significance at the 95% confidence level is indicated by **.
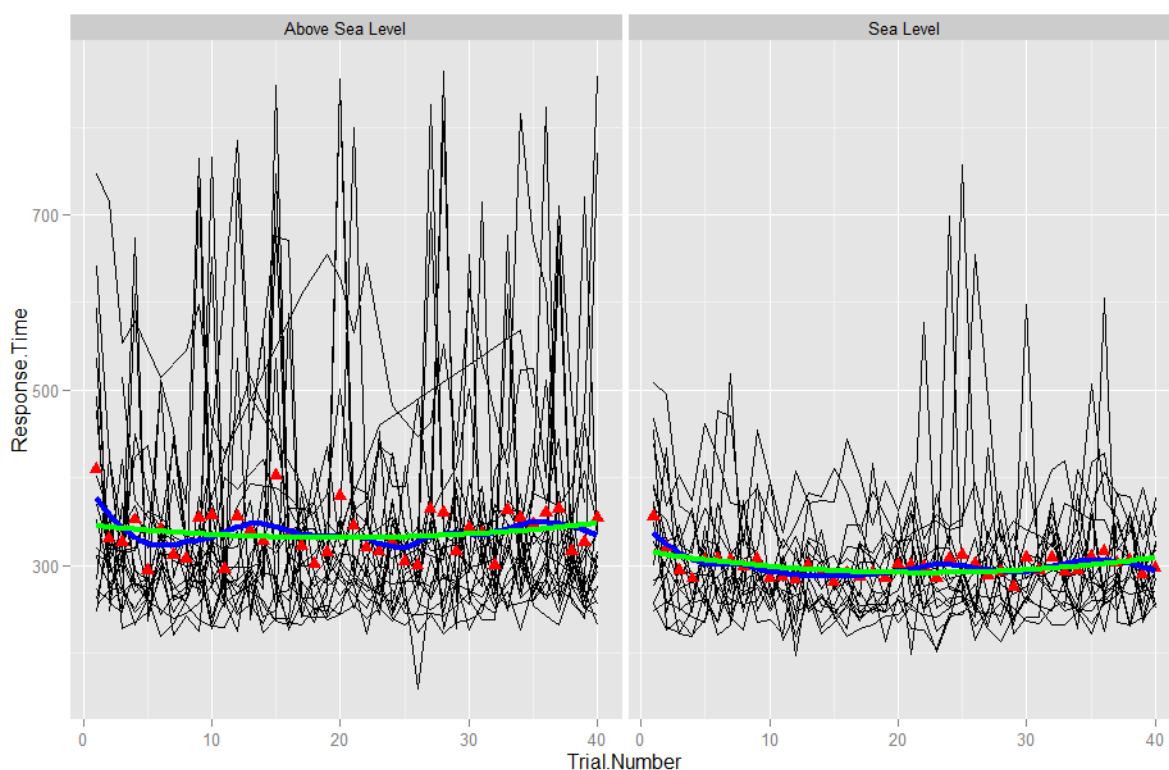

Figure 4: Trial Profiles by Altitude: Quadratic Linear Regression (Green) vs. Loess (Blue)

4.2   Linear Spline Regression

The loess curves in Figure 3 indicate localized curvature, particularly in the above sea level case. This feature makes linear spline regression a promising candidate for capturing the trend. Spline regression provides piecewise linear fits in which a set of separate linear models are fit in localized areas and joined together to estimate a curve. For ease of interpretation, we used truncated linear splines of degree one (as opposed to quadratic, cubic or B-splines). The basis of truncated linear splines is provided by using explanatory variables with the following form:

$$(\text{Trial Number} - \kappa_k)_+ = \begin{cases} \text{T}rialNumber - \kappa_k & \text{if Trial Number} > \kappa_k \\ 0 & \text{otherwise} \end{cases}$$

where $\kappa_1, \ldots, \kappa_K$ is a set of "knots" - points at which two separately sloped lines join together. The full linear spline regression equation is:

$$E(\text{Response Time}|\text{Trial Number}) = \beta_0 + \beta_1 * \text{Trial Number} + \sum_{k=1}^{K} \beta_{1+k} * (\text{Trial Number} - \kappa_k)_+ \quad (1)$$

Using the loess curves as a visual guide, the shape suggests that changes in slope occur about every 5 trials with some gaps. Accordingly, we chose knots at trial numbers 5, 15, 25, and 35. An advantage of the truncated linear spline is its ease of interpretation: in these models, the estimated coefficients of the spline variables are interpreted as the additional slope effect for a given trial range. To start, β_1 (the coefficient associated with trial number) is the slope from trial number 1 to trial number 5. Next, for example, the additional slope effect from trial number 6 to trial number 15 is β_2. The total slope from trial number 6 to trial number 15 is $\beta_1 + \beta_2$. Table 3 presents estimated coefficients, standard errors, and statistical significance from the linear spline fit to the altitude data.

Table 3: Linear Spline Regression Results by Altitude

	Above Sea Level	Sea Level
Covariate	Estimate (s.e.)	Estimate (s.e.)
Intercept (β_0)	394.3 (25.58)**	344.9 (14.42)**
Trial Number (β_1)	-16.14 (6.49)**	-9.35(3.66)**
Spline: Knot at 5 (β_2)	19.75 (7.76)**	8.41 (4.38)*
Spline: Knot at 15 (β_3)	-6.62 (3.31)**	1.85 (1.89)
Spline: Knot at 25 (β_4)	6.31 (3.25)*	-.38 (1.87)
Spline: Knot at 35 (β_5)	-6.45 (6.13)	-1.54 (3.55)

Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

For the sea level model, there is a statistically significant negative slope between trials 1 and 5 ($\hat{\beta}_1 = -9.35$). The magnitude of the slope is statistically significantly different and slightly negative from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2 = -.94$). The slope does not significantly change after trial 15. For the above sea level model, there is also a statistically significant negative slope between trials 1 and 5

$(\hat{\beta}_1 = -16.14)$. The magnitude of the slope is statistically significantly different and positive from trials 6 to 15 $(\hat{\beta}_1 + \hat{\beta}_2 = 3.61)$, and continues to significantly change until trial 35. It is negative from trials 16 to 25 $(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = -3.01)$, positive from trials 26 to 35 $(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 = 3.30)$, and does not significantly change after trial 35.

These results suggest possible differences in average response time patterns for subjects at different altitudes. Both altitudes exhibit a learning effect; the average response times show a downward slope in the first 5 or so trials; however, while the sea level subjects on average don't exhibit large changes after that time, the above sea level subjects on average revert back towards their pre-learning response times and don't maintain a constant effect after this point.

Results from the linear spline regression are consistent with patterns that were observed in the previous figures. Figure 5 shows that the linear spline regression captures the curvature of the loess curves in Figure 3.
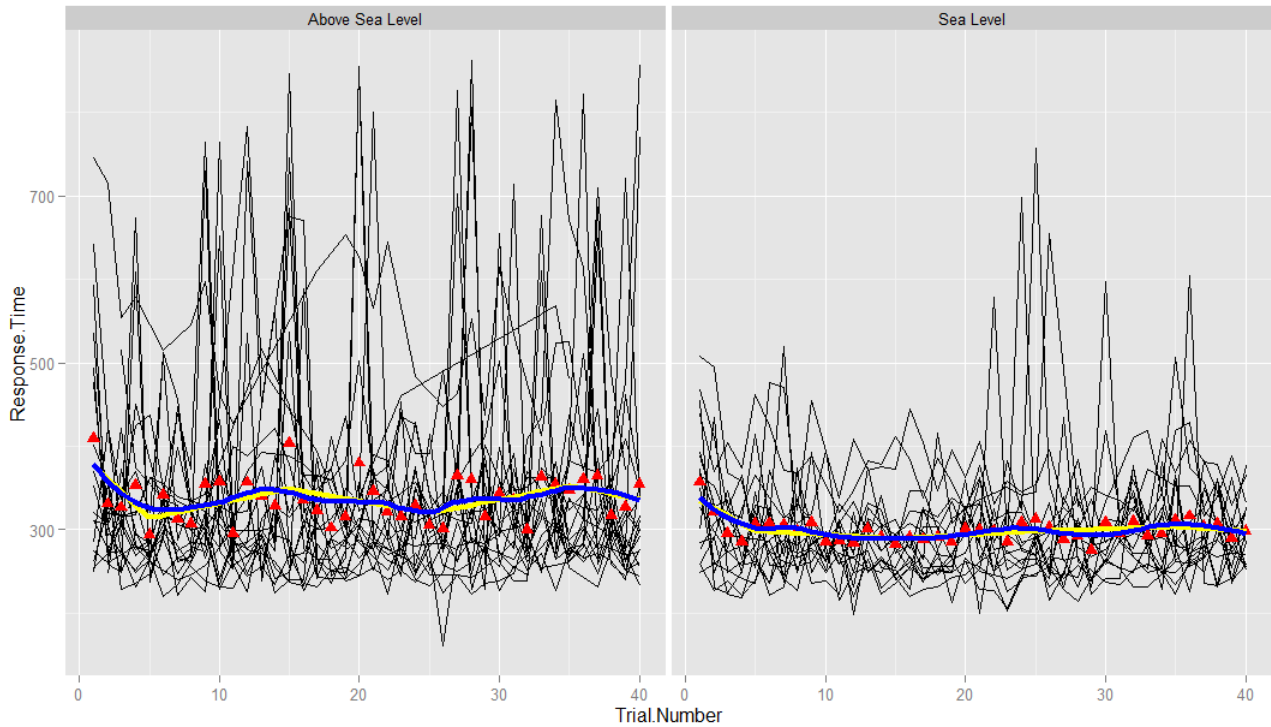

Figure 5: Trial Profiles by Altitude: Linear Regression with Splines (Yellow) vs. Loess (Blue)

We can formally test for differences in mean profiles for each altitude by fitting one model that has interaction terms for altitude level (see Table 4 for estimated coefficients, standard errors and statistical significance). In this model, only the change in slope from trials $16 - 25$ and $26 - 35$ are statistically significantly different between altitudes. This may be due to small sample size.

Table 4: Linear Spline Regression Results

Covariate	Estimate (s.e.)
Intercept	394.3 (20.90)**
Sea Level	-49.44 (30.40)
Trial Number	-16.14 (5.31)**
Sea Level * Trial Number	6.79 (7.72)
Spline: Knot at 5	19.75 (6.35)**
Spline: Knot at 15	-6.62 (2.71)**
Spline: Knot at 25	6.31 (2.66)**
Spline: Knot at 35	-6.45 (5.02)
Sea Level * Spline: Knot at 5	-11.34 (9.23)
Sea Level * Spline: Knot at 15	8.47 (3.96)**
Sea Level * Spline: Knot at 25	-6.69 (3.91)*
Sea Level * Spline: Knot at 35	4.91 (7.39)

Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

5 Linear Mixed Models

The analysis in Section 4 assumes that the observations are independent; however, independence is clearly not the case in repeated measures data. In these analyses, the correlation among response times within a subject needs to be taken into account. Linear mixed models are one tool that accounts for the correlation with a trial-constant subject-specific response time effect. In this analysis, we introduce only a random intercept into the linear model. This allows each subject's intercept (average response time) to be different from the others. Equation 1 can be extended to include a random intercept as follows:

$$E(\text{Response Time}|\text{Trial Number}) = \beta_0 + \beta_1 * \text{Trial Number} + \sum_{k=1}^{K} \beta_{1+k} * (\text{Trial Number} - \kappa_k)_+ + u_i \quad (2)$$

where i denotes the subject, j denotes the trial number and u_i is the random intercept. Table 5 displays results from incorporating a subject random intercept in the linear spline regression models from Section 4. Accounting for subject-level correlation results in smaller standard errors. Although there are some changes in the estimates of the coefficients, earlier conclusions concerning statistically significant effects are unchanged.

9

Table 5: Linear Spline Mixed Model Results

Covariate	Estimate (s.e.)
Intercept	394.4 (21.45)**
Sea Level	-49.78 (31.40)
Trial Number	-14.86 (4.64)**
Sea Level * Trial Number	5.65 (6.73)
Spline: Knot at 5	18.87 (5.54)**
Spline: Knot at 15	-7.48 (2.37)**
Spline: Knot at 25	6.47 (2.32)**
Spline: Knot at 35	-4.81 (4.38)
Sea Level * Spline: Knot at 5	-10.64 (8.05)
Sea Level * Spline: Knot at 15	9.37 (3.46)**
Sea Level * Spline: Knot at 25	-6.81 (3.41)**
Sea Level * Spline: Knot at 35	3.00 (6.45)
Random Intercept Variance	2640

Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

In addition to accounting for correlation, linear mixed models also provide tools for understanding components of variability in the data. The subject-level variance indicates how much of the variability in the data is due to inclusion of a subject random intercept. In this data set, we find that the subject effect accounts for about 28% of the overall variability. While the subject-specific intercept is capturing considerable variability in the data, the level of variability (Random Intercept Variance = 2640) is quite large. The magnitude of this variability suggests that inclusion of more subject-specific characteristics that explain subject-specific intercept shifts (age, gender or education level, for example) might result in a better model - one that reduces subject-level variability. Mixed models can also be used to assess individual effects with estimated subject-specific random intercepts; these quantities can potentially be used to identify outliers, and we propose to explore their use in Phase III of the study plan as a potential strategy for identifying "stressed" subjects.

6   Conclusions

This report presents exploratory data analyses of trial profiles of simple reaction time for sea level and above sea level subjects. A nonparametric scatterplot smoother (the loess) graphically depicted a difference in response time and variability in response time (as observed by the wiggliness of the curve) between sea level and above sea level subjects. These descriptive analyses suggested the utility of a linear spline model to describe average trial profiles. We observed that a linear spline analysis captured the shape of the average profiles and indicated statistically significant differences in the shape of the average profile for sea level and above sea level subjects. We did not observe differences in the conclusions regarding the shape of the average trial profile when we incorporated subject-level variability using linear mixed models; however we did see considerable variability in the subject-specific intercepts. Further investigation into the nature and utility of subject-specific random intercepts is well-suited to efforts aimed at identifying "stressed" subjects.

Simple Reaction Time Repeated Measures Analysis: Phase II Report

Contents

List of Figures

List of Tables

1 Statement of the Challenge

Although much has been published on summary statistics of multi-trial cognitive function tests, little has been done to leverage all the trial data upon which these summary measures are based. It may be possible to use these data in ways that can improve on current strategies to identify people with head injury, depression, PTSD, and age-related cognitive decline. A first step in this process is to identify robust methods that describe patterns in trial data such that "normal" individuals can be distinguished from those outside the normal range.

2 Data

This report presents repeated measures analysis of the "SRT - Air Force" data from the Excel file "Altitude and Air Force Trial by Trial Data(updated).xlsx" received by Resnick, Chodorow and Associates on April 21. This analysis makes some data assumptions and corrections:

- Lapsed and fast trials ("response = Lapse" or "response = Fast (Correct)') are excluded from the analysis.

- Subject L0709 is deleted - this subject's data is repeated twice with slighly different data.

After these exclusions, we are left with a dataset of $6,327$ observations uniquely identified by ID, trial number, and condition. We focus our attention on these three variables in addition to reaction time, the primary variable of interest. Table 1 presents summary information about the analysis data set.

Table 1: Summary Information about the Air Force Data

	Concussed	Healthy
Number of Subjects	6	153
Average Number of Trials Per Person	39.3	39.8
Range of Number of Trials Per Person	36-40	36-40
Average Response Time	314.8	310.7
Range of Response Time	180-832	181-891
Average of Subject-Average Response Time	317.5	311.0
Range of Subject-Average Response Time	222.3-480.4	230.1-554.2

3 Exploratory Data Analysis

The Phase I report included a descriptive assessment of the response time trial profiles for the altitude data. We repeat the same analyses for the Air Force data to compare characteristics of the two datasets.

3.1 Air Force Data

Figure 1 plots the trial profiles of response time for each subject by condition (concussion vs. healthy). Each line represents one subject's response times traced over up to 40 trials. The trial

profiles appear erratic; within condition, there is no obvious visual trend that is common across subjects. It is particularly difficult to visually distinguish differences between groups because of the difference in the number of subjects for the two conditions: only 6 concussed subjects versus 153 healthy subjects.
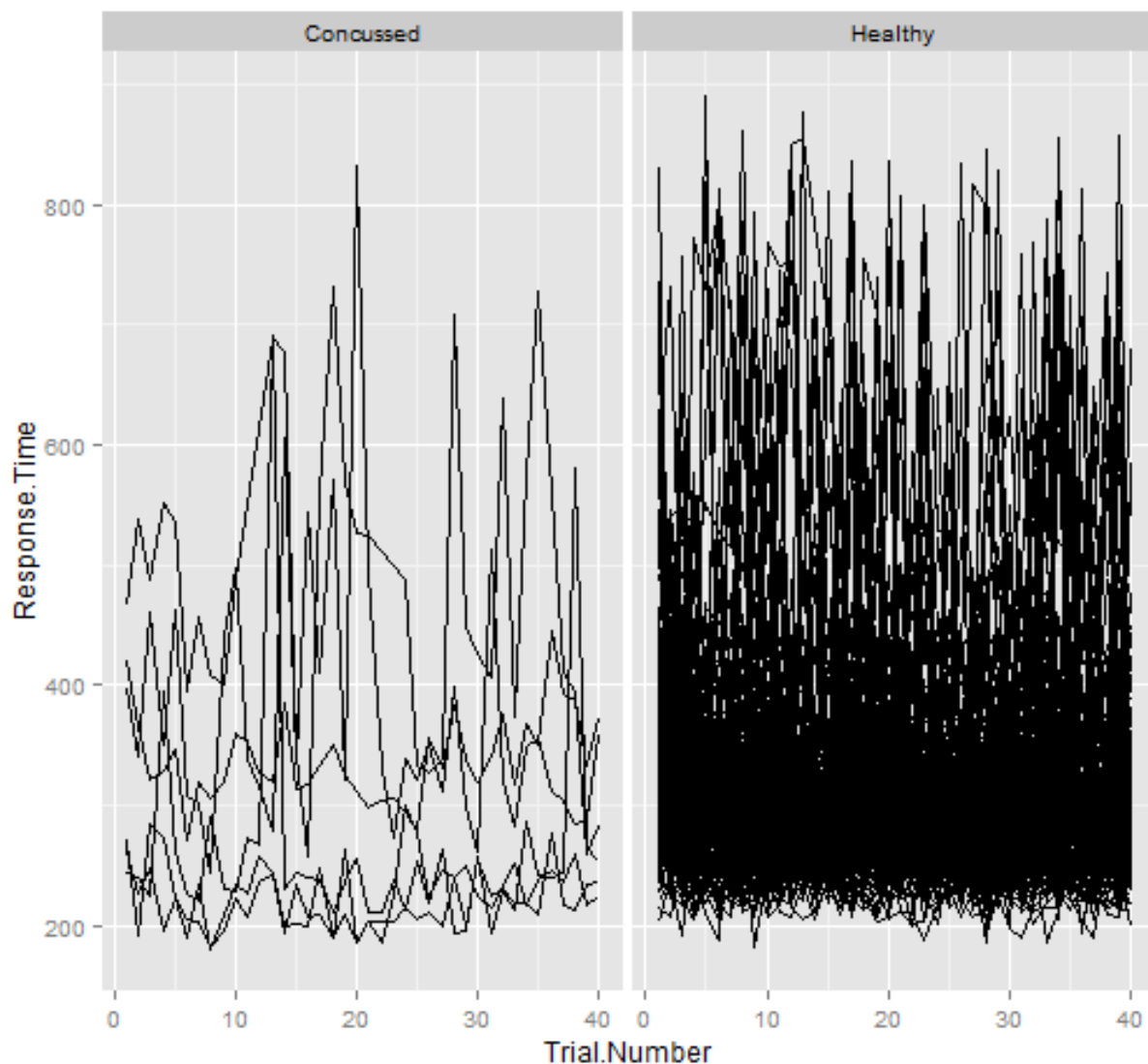

Figure 1: Trial Profiles of Response Time by Condition

Figure 2 adds average response times by trial number (red triangles) to the trial profiles shown in Figure 1. The triangles depict a stable trend line for trial averages for healthy subjects. We observe more erratic averages in the concussed subjects, but this may be due to the small sample size.
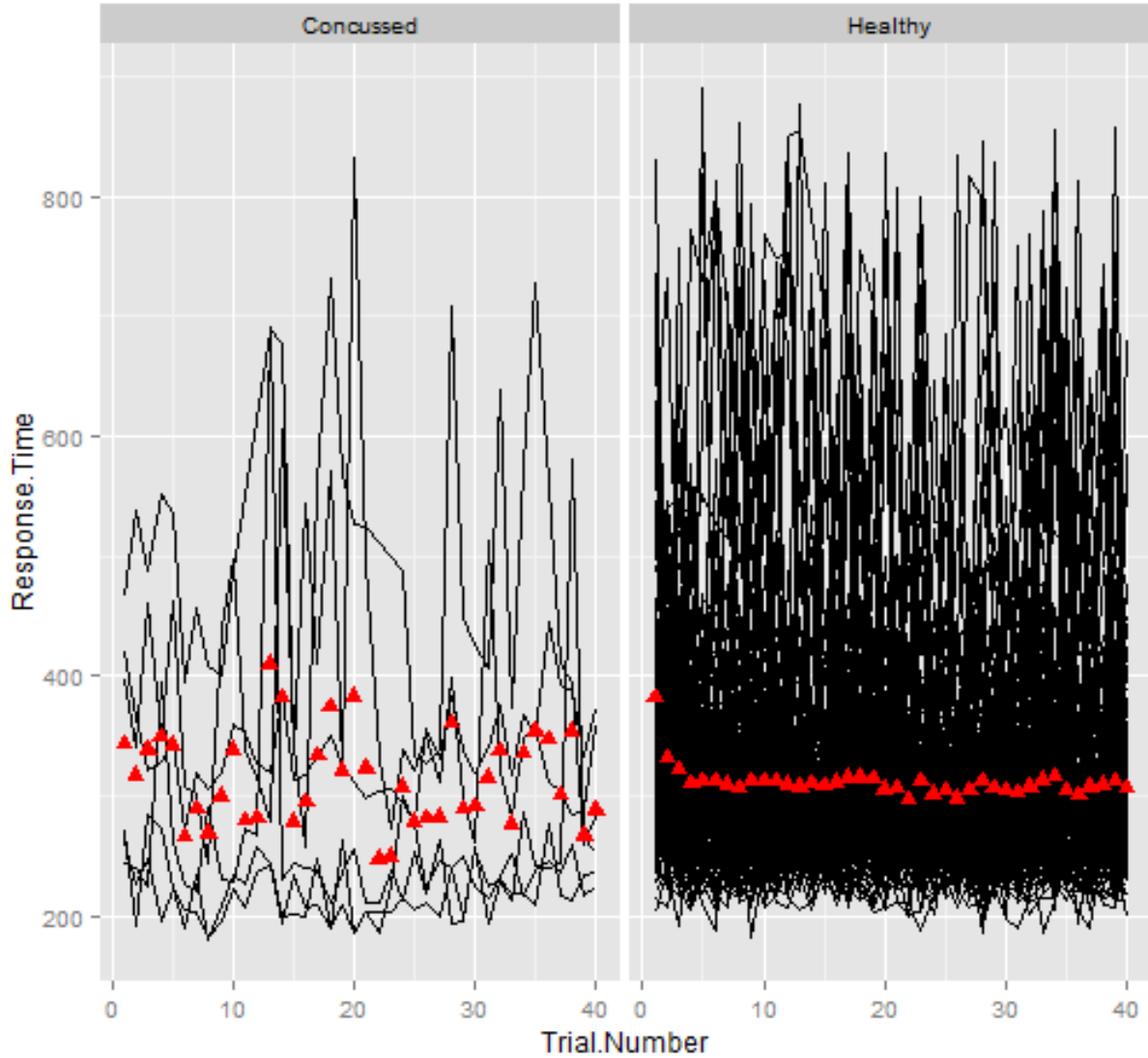
3

Figure 2: Trial Profiles by Condition: Trial Averages

Figure 3 includes a loess curve fit to the trial data. Please see the Phase I report for details about this method. Imposing loess curves on the plots of trial profiles facilitates visualization of differences in their means and variability: the loess curve for the healthy subjects is stable and consistently lies just above 300 with the exception of the first few trials. In contrast, the loess curve for the concussed subjects exhibits a very wiggly behavior that goes as low as about 280, and as high as about 340. Although there is more curvature in the trial averages for the concussed subjects, both healthy and concussed subjects have a similar average of trial-level average response times (the red triangles): 311 and 314, respectively. In addition to concussed subjects having more curvature in their average trial profiles, the trial-level response times are more variable around the

4

loess curve for the concussed subjects than for the healthy subjects.

The loess curves provide insight into appropriate models for average response times across trials. The loess curve for the concussed subjects is very wiggly. Notably, the curve exhibits pronounced humps that arise from changes in slope roughly around trials 5, 15, 20, 25 and 35. The healthy subjects' average trial profile is strikingly linear after about trial 5.
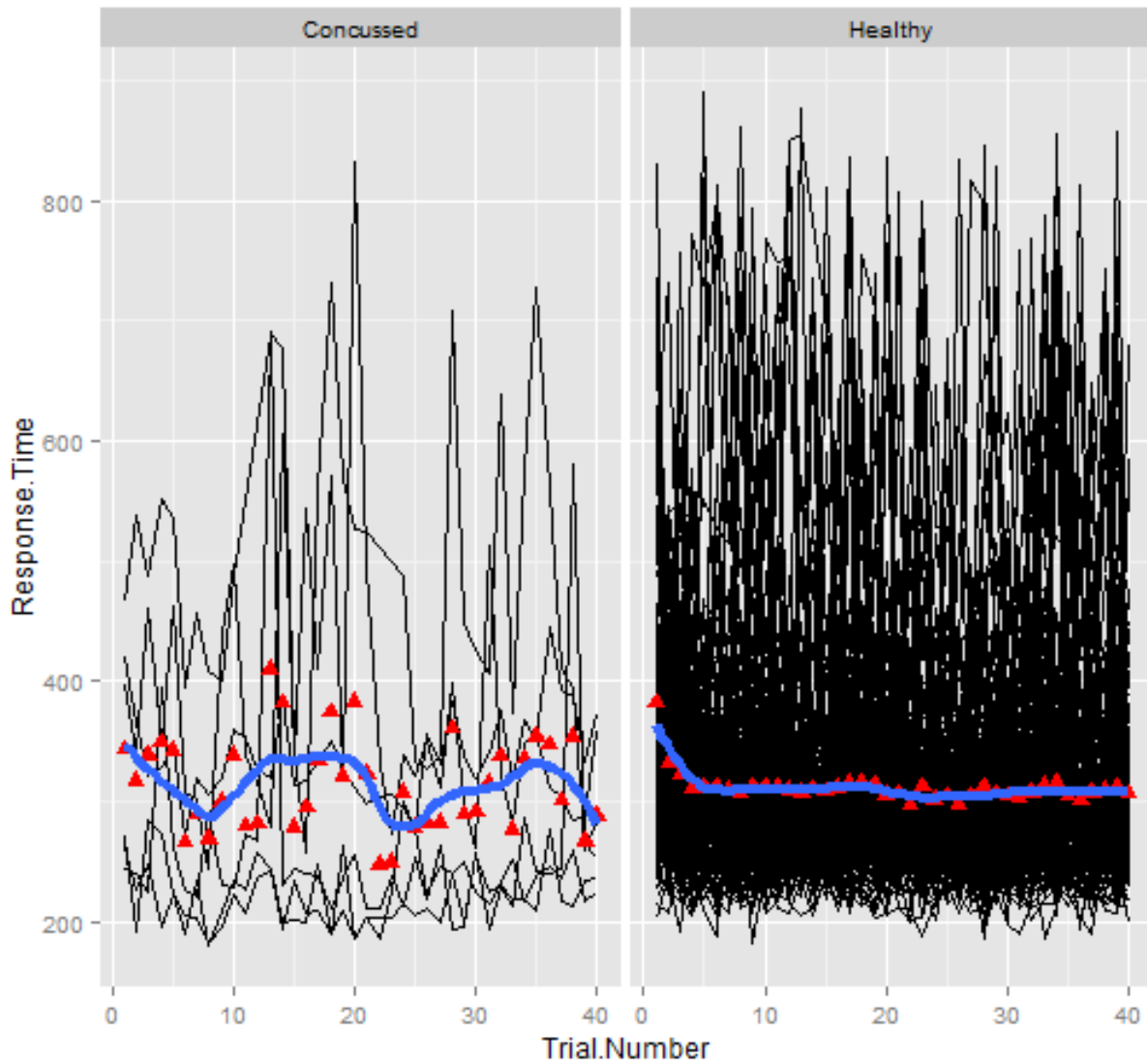

Figure 3: Trial Profiles by Condition: Loess

3.2 Comparison with Altitude Data

There are important differences in the nature of the altitude and the Air Force data to keep in mind. The altitude data have a comparable number of subjects for both the normal and non-normal condition groups, while the number of subjects is very unbalanced in the Air Force data. Nonetheless, the average trial profiles for normal subjects are very similar in both the altitude and the Air Force data: there is a dip in average response time until about trial 5, followed by a linear and flat average response time. In addition, the average trial-level response times for normal subjects are around 300 for both datasets. Average trial profiles for the non-normal subjects are more variable than for normal subjects in both datasets, but the non-normal average trial profile is a lot more variable for the Air Force data than for the altitude data, possibly due to the small sample size. While the loess curve for non-normal subjects in the Air Force data exhibits a lot more curvature than the non-normal subjects in the altitude data (particularly after trial 10), the pivot points (the places where the slopes change) in the shape of the curve appear roughly similar to those observed in the altitude data.

4 Linear Models (Linear Spline Regression)

In Phase I of this work, we identified a reasonable model for response time trial profiles in the altitude data. Linear spline regression with knots at trial numbers 5, 15, 25, and 35 provided an adequate approximation of the shape of the average trial profiles. In this section we apply this model to the Air Force data.

4.1 Air Force Data

The loess curves in Figure 3 indicate localized curvature, particularly for concussed subjects. Just as in the altitude data, this feature makes linear spline regression a promising candidate for capturing the trend. Please see the Phase I report for details about the specifications of the linear spline regression that was fit to the altitude data; this model is now used for the Air Force data.

Table 2: Linear Spline Regression Results by Condition

| | Concussed | Healthy |
Covariate	Estimate (s.e.)	Estimate (s.e.)
Intercept (β_0)	360.2 (48.69)**	374.9 (6.90)**
Trial Number (β_1)	-12.82 (12.35)	-13.95 (1.75)**
Spline: Knot at 5 (β_2)	17.28 (14.77)	14.68 (2.09)**
Spline: Knot at 15 (β_3)	-10.14 (6.40)	-1.70 (0.90)*
Spline: Knot at 25 (β_4)	11.08 (6.37)*	1.54 (0.89)*
Spline: Knot at 35 (β_5)	-15.71 (11.96)	-0.77 (1.69)

Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

Table 2 presents the estimates of the linear spline model for the Air Force data. For the model of healthy subjects, there is a statistically significant negative slope between trials 1 and

6

5 ($\hat{\beta}_1 = -13.95$). The magnitude of the slope is statistically significantly different and slightly positive from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2 = .73$). The slope significantly changes from trials 16 to 25 and 26 to 35, but these changes are small (-1.70 and 1.54, respectively). For the concussed model, only the knot at 25 is estimated to reflect a statistically significant change in slope; however the magnitude of all the estimates of the changes in slope are quite large (a range of about -16 to 17).

These results suggest possible differences in average response time patterns for subjects with different conditions. Both conditions exhibit a learning effect; the average response times show a downward slope in the first 5 or so trials; however, while the healthy subjects on average don't exhibit large changes after that time, the concussed subjects on average revert back towards their pre-learning response times and don't maintain a constant effect after this point. While these changes in the average trial profiles for concussed subjects are not statistically significant, they are estimated to be quite large.
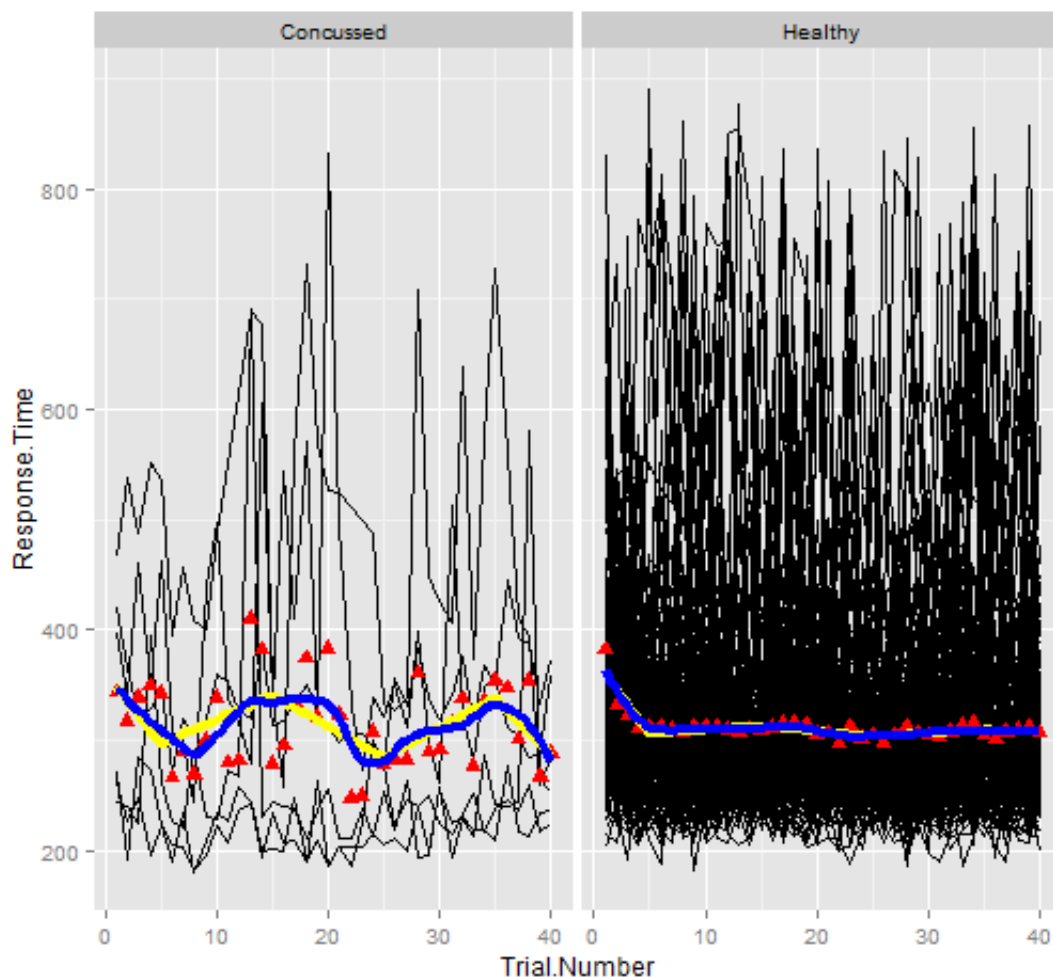

Figure 4: Trial Profiles by Condition: Linear Regression with Splines (Yellow) vs. Loess (Blue)

7

Results from the linear spline regression are consistent with patterns that were observed in the previous figures. Figure 4 shows how the linear spline regression captures the curvature of the loess curves in Figure 3.

4.2 Comparison with Altitude Data

We observe some notable differences in the coefficient estimates for the linear spline model fit to the altitude and the Air Force data. For normal subjects, we find the same decrease in average response time between trials 1 and 5, with a larger estimated negative slope for the Air Force data (-13.95 vs -9.35). Changes in slope from trial 6 to 10 are estimated in both datasets, however this change in slope results in a slightly negative slope for the altitude data and a slightly positive slope in the Air Force data. After trial 10, significant changes in slope are estimated in the Air Force data, but not the altitude data. However, the significant changes in slope in the Air Force data are small in magnitude (of comparable levels to those estimated in the altitude data). It is important to keep in mind that there are 153 normal subjects in the Air Force data, but only 17 normal subjects in the altitude data.

For the non-normal subjects, all changes in slope for the altitude dataset are statistically significant (except for the knot at 35), while only one change in slope is statistically significant in Air Force dataset (the knot at 25). The magnitude of the estimates of the slope from trial 1 to 5 and the change in slope for trial 6 to 15 are comparable in the Air Force and altitude datasets (-12.82 vs. -16.14 and 17.28 vs. 19.75), but the estimated slopes beyond trial 15 are larger in the Air Force data than in the altitude data. This result reflects the more pronounced curvature in the average trial profiles observed for concussed subjects in the Air Force data.

5 Linear Spline Mixed Model

The analysis in Section 4 assumes that the observations are independent; however, independence is clearly not the case in repeated measures data. In these analyses, the correlation among response times within a subject needs to be taken into account. Linear mixed models are one tool that accounts for the correlation with a trial-constant subject-specific response time effect. Please see the Phase I report for details about the specification of the linear spline mixed model we fit to the altitude data, which we now fit to the Air Force data.

5.1 Air Force Data

For the separate models of healthy and concussed subjects, the estimates are very similar to those in Table 2 which do not incorporate the subject random effect. The main difference is that the standard errors on the trial number and spline coefficients are smaller because some of the variability is captured by allowing subject-level variability via the random effects. As a result, all changes in slope (i.e. all the coefficients on the linear spline functions) are statistically significant for the concussed subject. The conclusions for the average healthy trial profiles remains unchanged (i.e. the statistical significance remains the same but at a stricter level of confidence).

Table 3: Linear Spline Mixed Model Results

Covariate	Concussed Estimate (s.e.)	Healthy Estimate (s.e.)	Combined Estimate (s.e.)
Intercept	358.7 (52.41)**	374.7 (6.97)**	358.80 (35.74)**
Healthy			15.91 (36.44)
Trial Number	-12.07 (8.27)	-13.90 (1.49)**	-12.11 (7.54)
Healthy * Trial Number			-1.80 (7.69)
Spline: Knot at 5	16.87 (9.89)*	14.68 (1.79)**	16.89 (9.02)*
Spline: Knot at 15	-10.50 (4.28)**	-1.74 (0.77)**	-10.48 (3.91)**
Spline: Knot at 25	10.40 (4.27)**	1.54 (0.76)**	10.43 (3.89)**
Spline: Knot at 35	-14.51 (8.01)*	-0.74 (1.44)	-14.56 (7.31)**
Healthy * Spline: Knot at 5			-2.22 (9.20)
Healthy * Spline: Knot at 15			8.74 (3.98)**
Healthy * Spline: Knot at 25			-8.89 (3.97)**
Healthy * Spline: Knot at 35			13.82 (7.45)*
Random Intercept Variance	10106	2106	2360
Residual Variance	6761	5585	5628

Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

We can formally test for differences in mean profiles by fitting a model that has interaction terms for each condition level. The results from this combined model are in Table 3. In this model, the change in slope for trials $16-25$, $26-35$ and $35-40$ are statistically significantly different for healthy versus concussed subjects with estimates of 8.74, -8.89, and 13.82, respectively. That is, the model picks up that the average trial profile for concussed subjects has a different shape after trial 15 than the healthy subjects. For example, for concussed subjects the slope from trial 6 to 10 and trial 16 to 25 is 4.78 $(-12.11 + 16.89)$ and -5.70 $(-12.11 + 16.89 - 10.48)$, respectively. This change in slope, -10.48, is statistically significant. For healthy subjects, the slope from trial 6 to 10 and trial 16 to 25 is 0.76 $(-12.11 - 1.80 + 16.89 - 2.22)$ and $-.98$ $(-12.11 - 1.80 + 16.89 - 2.22 - 10.48 + 8.74)$, respectively. This change in slope, -1.74, is statistically significantly different (and smaller) than the change in slope for the concussed subjects (-10.48).

5.2 Comparison with Altitude Data

The results from the fully interacted linear spline mixed model are similar for the altitude and the Air Force data. Some main similarities and differences are:

- *Non-Normal Profiles*: In both datasets we find statistically significant differences in the change of slope at knot points 5, 15 and 25 for non-normal subjects; the direction of the changes is the same, but the magnitude of the changes are larger in the Air Force data. In the Air Force data only, the coefficient for the difference in the change of slope for the non-normal subjects at a knot of 35 is also statistically significant. Results about the spline coefficients for non-normal subjects reflect the curvature that is observed in the average trial profiles for non-normal subjects, which is more pronounced in the Air Force data.

- *Normal Profiles*: In both datasets we find statistically significant differences in the change of slope at knot points 5 for normal subjects; the direction of the changes is the same, but the magnitude of the changes are larger in the Air Force data. Also, in both datasets we find that the change in slope at knot points beyond trial 5 are of small magnitude even though some are statistically significant in the Air Force data. Results about the spline coefficients for normal subjects reflect the linearity that is observed in the average trial profiles for normal subjects.

- *Difference Between Normal and Non-Normal Profiles*: In both datasets we find statistically significant differences in the change of slope between normal and non-normal subjects at knot points 15 and 25; the direction of the change is the same and the magnitude of the differences are comparable. In the Air Force data only, the coefficient for the difference in the spline function between normal and non-normal subjects at knot 35 is also statistically significant. Results about the coefficients associated with the interaction of condition and spline functions reflect the differences in the shape of the curve between average trial profiles for normal and non-normal subjects: a flat, linear trend for normal subjects and a wiggly trend for non-normal subjects.

- *Learning Effect*: The slope of the trial profile between trials 1 and 5 is negatively sloped (indicative of a learning effect) for both datasets, but this decrease in response time is only statistically significant in the altitude data.

6   Conclusions

This report presents exploratory analysis and modeling results of concussed and healthy subjects in the Air Force data. The linear spline mixed model of average trial profiles developed in Phase I of this series of reports is applied to the Air Force data. We find many commonalities between the average trial profiles in the altitude and Air Force datasets. These commonalities are evident in the plots of trial profiles as well as in the estimated coefficients in the model. Most notably, we find evidence of a learning effect depicted as a downward slope in response time for the first few trials, a constant linear average response time trial profile for normal subjects, and a variable and wiggly average response time trial profile for non-normal subjects.

   The differences in the shape of the average response time trial profile suggest the possibility of a tool for distinguishing non-normal subjects from normal subjects. While the typical subject's trial profile is somewhat erratic, it may be possible to implement a smoother at the subject level to pick up any distinct shape in the profile as compared to the average trial profiles. Such techniques are common in functional data analysis. Mixed models also offer a possibility for distinguishing non-normal subjects from normal subjects. The random intercept model implemented in this report produces estimates of subject-specific shifts from the average response time. It may be useful to extend the mixed model to incorporate random slopes, which produce subject-specific estimates of changes in slope from the average changes in slope. Further investigation into the utility of subject-specific random intercepts/slopes and subject-level smoothers is well-suited to efforts aimed at identifying non-normal subjects.

Repeated Measures Analysis: Phase III Report

1 Statement of the Challenge

Although much has been published on summary statistics of multi-trial cognitive function tests, little has been done to leverage all the trial data upon which these summary measures are based. It may be possible to use these data in ways that can improve on current strategies to identify people with head injury, depression, PTSD, and age-related cognitive decline. Previous work has identified viable models for describing patterns in trial data. The next step in this research is to explore how these models can be used to (1) extend inferences from a single SRT assessment to two SRT assessments and (2) extend inferences to DANA's code sub-learning assessment. The overarching objective of this work is to capitalize on DANA's repeated measures to establish a robust method for identification of conditions such as concussion, depression, dementia, sleep deprivation, and PTSD.

2 Simple Reaction Time Assessments: Full Set of 80 Trials

With a focus on simple reaction time (SRT), this section presents analysis of the "Altitude data set" that extends previous findings concerning the shapes and slopes of SRT curves for individuals at sea level and at altitude. In our earlier work, we showed that the shape of participants' SRT curves differed significantly depending on whether they were at sea level or at altitude. An appealing feature of those findings was the participants acted as their own controls. In these analyses, we used the first of two SRT administrations (SRT1). In this section of this Phase III report, we explore whether analysis of SRT2 adds useful information to our previous findings. The idea that SRT2 may add information is based on previous work indicating that these additional data points provided insight into individual-level performance metrics. It is therefore possible that adding more data points will make it easier for this strategy to identify and distinguish people with normal reaction time from those outside the normal range. Accordingly, we will use both SRT administrations from the "Altitude data set" by adding the second set of 40 SRT2 trials to the existing 40 SRT1 trials. All 80 trials will be examined together using the methods that were developed in earlier phases of this work.

2.1 Data

This section presents repeated measures analysis of the "SRT – Altitude" data from the Excel file "Altitude and Air Force Trial by Trial Data(updated).xlsx" received by Resnick, Chodorow and Associates on April 21, 2015. This analysis makes some simplifying data assumptions:

- Lapsed and fast trials ("response = Lapse" or "response = Fast (Correct)") are excluded from the analysis.

- Only trials at 5260m above sea level ("altitude = 1") and at sea level ("altitude = 3") are included in the analysis.

After these exclusions we are left with a dataset of 2,909 SRT trials. Note that data from both SRT1 and SRT2 are included in this analysis, resulting in a maximum of 80 trials per subject. We append trials from SRT2 to the trials from SRT1 to obtain a continuous set of 80 trials for each individual. This results in a dataset that is uniquely identified by ID, trial number, and altitude. We focus our attention on these three variables in addition to reaction time, the primary variable of interest. Table 2.1 presents summary information about the analysis data set.

Table 2.1: Summary Information about the Altitude Data

	Above Sea Level	Sea Level
Number of Subjects	21	17
Average Number of Trials Per Person	74.2	79.4
Range of Number of Trials Per Person	40*-80	75-80
Average Response Time	354.0	306.9
Range of Response Time	160-893	198-816
Average of Subject-Average Response Time	359.3	307.2
Range of Subject-Average Response Time	279.1-266.1	266.1-387.7

* The minimum of 40 trials comes from a subject with 10 trials in administration 1 and 30 trials in administration 2.

2.2   Exploratory Data Analysis

Our Phase I report assessed the SRT response time trial profiles for the altitude data for the first 40 SRT trials (SRT1). We repeat a similar analysis for the SRT data for all 80 trials.

Figure 2.1 plots the trial profiles of response time for each subject by altitude. Each line represents one subject's response times traced over up to 80 trials. The trial profiles appear erratic in that within altitude, there is no obvious visual trend that is common across subjects. While individual profiles are extremely difficult to distinguish in Figure 2.1, some average features are more apparent across the two altitudes: (1) above sea level response time measurements appear more variable, and (2) the bulk of the above sea level measurements seem to lie above the bulk of the sea level measurements.
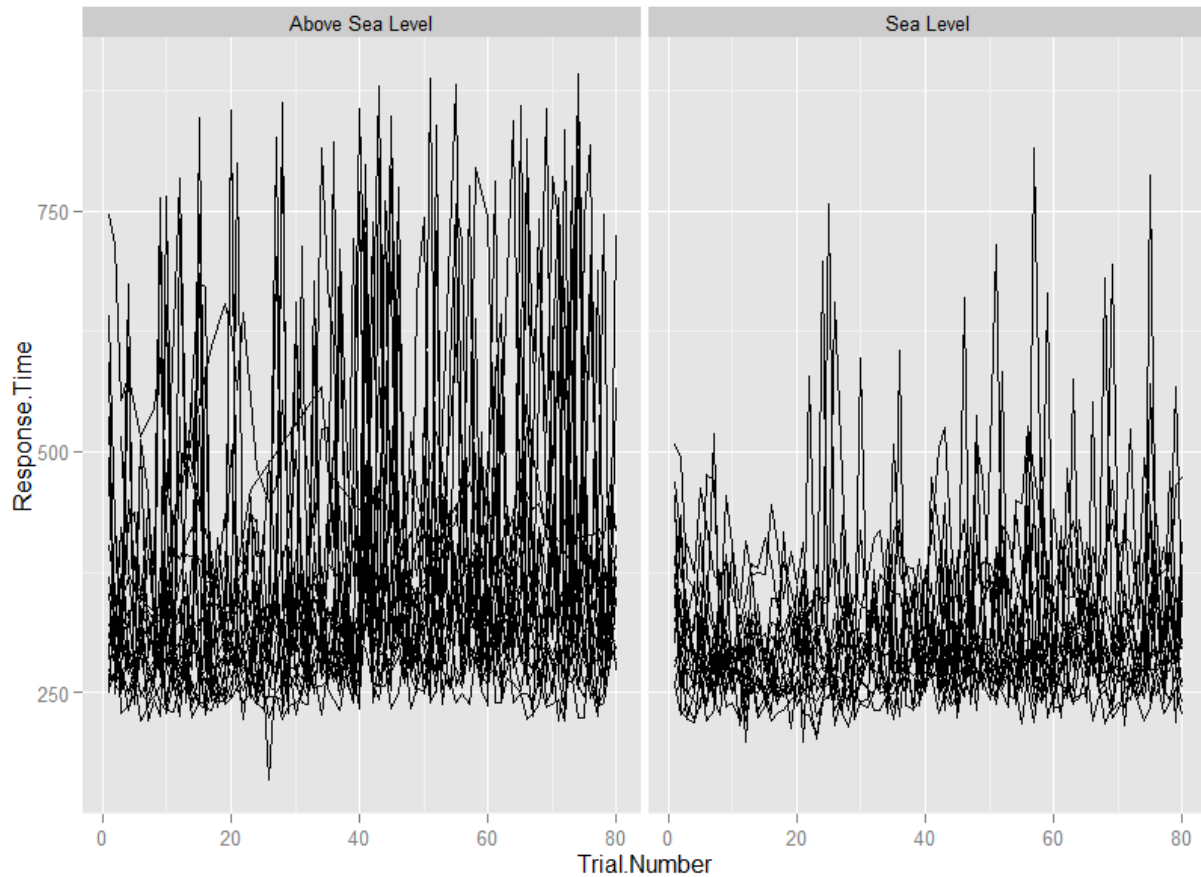
Figure 2.1: Trial Profiles of Response Time by Altitude

Figure 2.2 adds average response times by trial number (red triangles) to the trial profiles shown in Figure 2.1. The triangles confirm that the sea level trial averages generally lie below the above sea level averages, with a more pronounced difference in the second 40 trials. Additionally, the sea level trial averages form a relatively stable trend line, but show more variability in the trial averages in the second 40 trials. This contrasts with the more erratic averages that are observed in the above sea level subjects, which appear most variable in the first 40 trials. For both altitudes, the second set of 40 trials appear to have higher average response times – with the appearance of an upward slope around the transition from SRT1 to SRT2 trials.
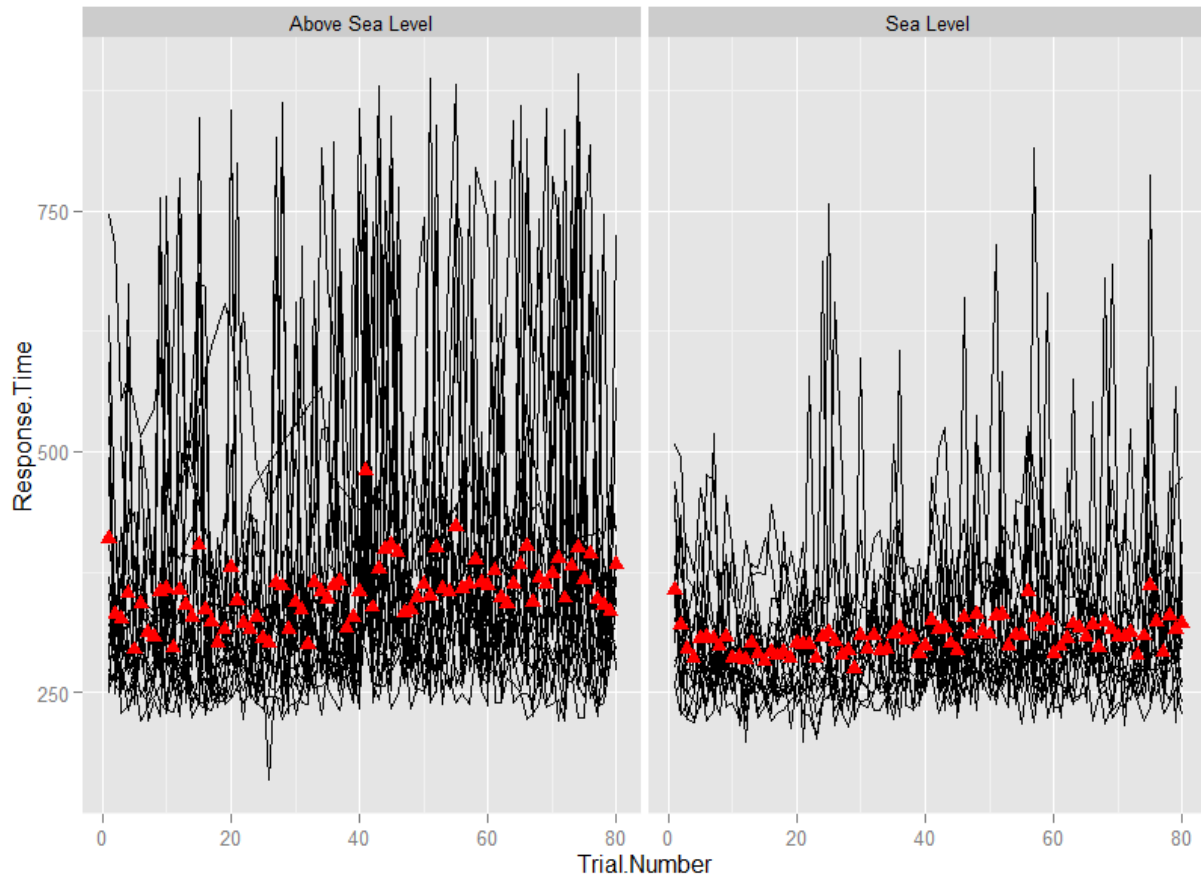
Figure 2.2: Trial Profiles by Altitude: Trial Averages

Figure 2.3 includes a loess curve fit to the trial data. Please see the Phase I report for details about this method. Imposing loess curves on the plots of trial profiles facilitates visualization of differences in their means and variability: the loess curve for sea level subjects hovers around 300 for the first 40 trials and slightly higher for the second 40 trials, while the loess curve for the above sea level subjects exhibits wigglier behavior that is frequently larger than 300 (reaching close to 400 around the transition from SRT1 to SRT2 data). In addition, average response times by trial (the red triangles) are more variable around the loess curve for the above sea level subjects than for the sea level subjects. Also, above sea level subjects exhibit a visually significant hump around trial 40 – the transition from SRT1 to SRT2 data.
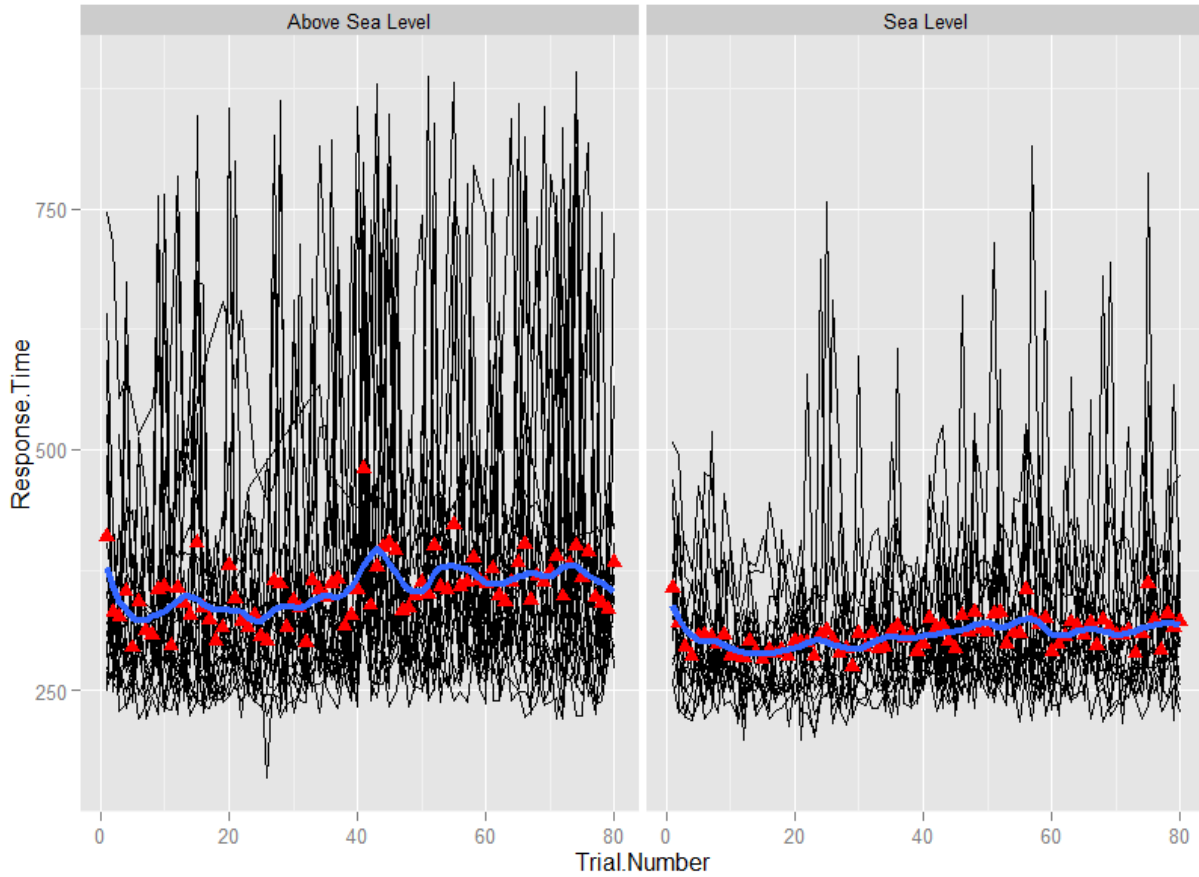
Figure 2.3: Trial Profiles by Altitude: Loess

2.3   Linear Models

In Phase I of this work, we identified a reasonable model for response time trial profiles in the altitude data. Linear spline regression with knots at trial numbers 5, 15, 25, and 35 provided an adequate approximation of the shape of the average trial profiles in the SRT1 data. In this section we apply the same model to the full set of 80 trials in the altitude data, with the addition of knot points for the second 40 trials. Using the loess curves as a visual guide, the shape of the second 40 trials suggests that changes in slope occur about every 5-10 trials. Accordingly, we chose knots at trial numbers 45,50,55,65, and 75. These correspond to the 5th, 10th, 15th, 25th and 35th trials of SRT2. The positions of these knot points for the SRT2 data differ from the SRT1 data, partly because of the transition at trial 40 from SRT1 to SRT2.

2.3.1   Linear Spline Regression

5

The loess curves in Figure 2.3 indicate localized curvature, particularly for above sea level subjects. This feature makes linear spline regression a promising candidate for capturing the trend. Please see the Phase I report for details about the specifications of the linear spline regression that was fit to the first 40 trials of the altitude data; this model is now used for the full set of 80 trials in the altitude data.

Table 2.2: Linear Spline Regression Results by Altitude

| | Above Sea Level | Sea Level |
Covariate	Estimate (s.e.)	Estimate (s.e.)
Intercept (β_0)	394.2 (27.76)**	344.9 (16.21)**
Trial Number (β_1)	-16.09(7.05)**	-9.34 (4.11)**
Spline: Knot at 5 (β_2)	19.63 (8.42)**	8.38 (4.92)*
Spline: Knot at 15 (β_3)	-6.35 (3.59)*	1.91 (2.12)
Spline: Knot at 25 (β_4)	5.35 (3.37)	-0.60 (2.01)
Spline: Knot at 35 (β_5)	1.92 (3.49)	0.55 (2.09)
Spline: Knot at 45 (β_6)	-13.61 (6.05)**	0.82 (3.58)
Spline: Knot at 50 (β_7)	16.31 (8.45)*	-1.42 (4.96)
Spline: Knot at 55 (β_8)	-9.07 (6.08)	-1.70 (3.59)
Spline: Knot at 65 (β_9)	3.59 (3.66)	2.44 (2.20)
Spline: Knot at 75 (β_{10})	-7.68 (6.75)	-1.04 (4.00)

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

Results in Table 2.2 indicate that for the sea level model, there is a statistically significant negative slope between trials 1 and 5 ($\hat{\beta}_1$ = -9.34). The magnitude of the slope is statistically significantly different and slightly negative from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2$ = -.96). In the sea level model, the slope does not significantly change after trial 15, nor are there statistically significant changes in the average trial response for the SRT2 data. The lack of significance in the sea level SRT2 data may be due to greater variability about the average response time profile in the SRT2 data; this is reflected in the larger standard errors in the knot coefficients for knots 45-75.

For the above sea level model, there is also a statistically significant negative slope between trials 1 and 5 ($\hat{\beta}_1$ = -16.09). The magnitude of the slope is statistically significantly different and positive from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2$ = 3.54), and statistically significantly different and negative from trials 16 to 25 ($\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ = -2.81). The slope of the above sea level model does not significantly change again until trial 45. Above sea level results for the first 40 trials are consistent with findings from the Phase I analysis, with the exception of statistical significance and magnitude of knot coefficients at the end of the 40 trial sequence (i.e. the knots at 25 and 35). Because the trend of the second set of trials begins at a relatively high average response time, the wiggliness at the tail end of the SRT1 trend is masked because the average trial profile is pulled upwards. This is also reflected in the difference in sign for the coefficient on the last knot point of the first set of 40 trials (knot at 35), which is positive ($\hat{\beta}_5$ =1.92) in this analysis but was negative ($\hat{\beta}_5$ =-6.45) in earlier analyses of only SRT1 data. However, neither of these coefficients is statistically significant. Regarding the second half of the average response time profile for above sea level subjects, there is a statistically significant negative slope between trials 46 and 50 ($\hat{\beta}_1 + \cdots + \hat{\beta}_6$ = -9.16) that significantly changes in a positive direction between trials 51 and 55($\hat{\beta}_1 + \cdots + \hat{\beta}_7$ = 7.15), and does not significantly change after trial 55.

6

These results suggest possible differences in average response time patterns for subjects at different altitudes. Both altitudes exhibit a learning effect; the average response times show a downward slope in the first 5 or so trials; however, while the sea level subjects on average don't exhibit large changes after that time, the above sea level subjects on average revert back towards their pre-learning response times until the transition point to SRT2. For these subjects, the average response time at the beginning of SRT2 mirrors the average response time at the beginning of SRT1. A similar learning effect is observed, followed by a minor reversion back towards their pre-learning response times. The average response time does not maintain a constant effect following the observed learning effects. The sea level subjects do not exhibit this second learning effect, although they do exhibit an average increase in response time after the transition to SRT2.

Results from the linear spline regression are consistent with patterns that were observed in the previous figures. Figure 2.4 shows that the linear spline regression captures the curvature of the loess curves in Figure 2.4.
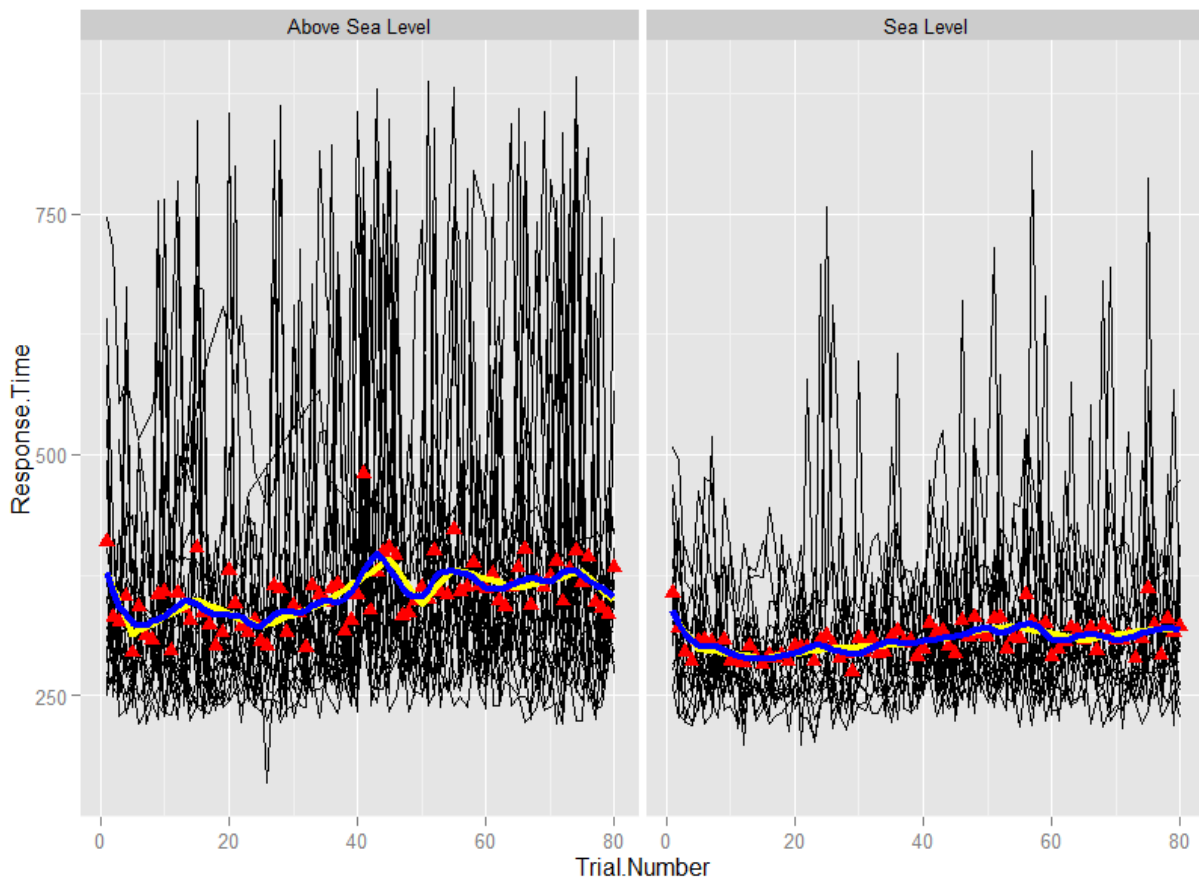

Figure 2.4: Trial Profiles by Altitude: Linear Regression with Splines (Yellow) vs. Loess (Blue)

Figure 2.5 presents the estimated linear spline fit for above sea level and sea level subjects in the Altitude data. The difference in scale, as compared to Figure 4, exaggerates the wiggliness while

7

illustrating the estimated increase in average response time around the transition from SRT1 to SRT2 data. This is followed by a steep decrease for the above sea level subjects contrasted with the more gradual decrease for the sea level subjects.
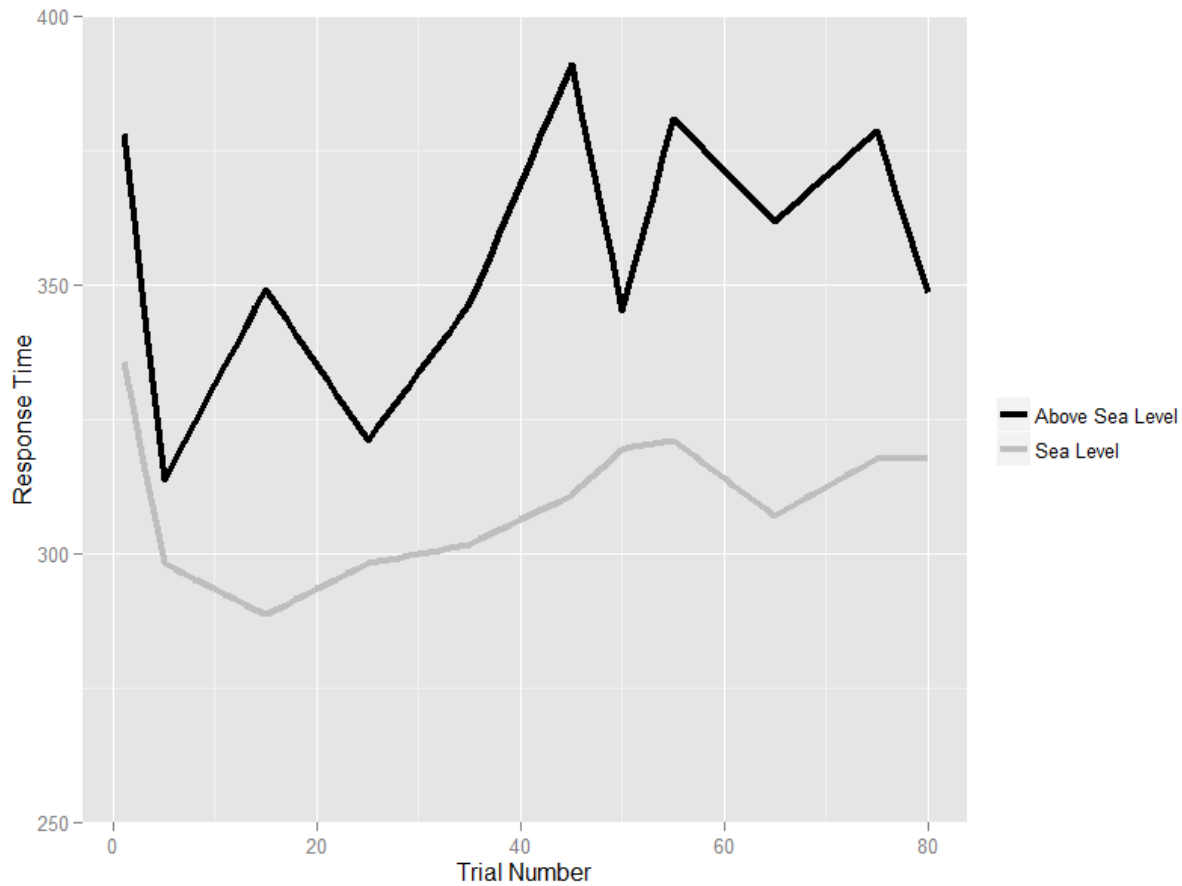

Figure 2.5: Linear Spline Fit of Response Time by Altitude

2.3.2 Linear Spline Mixed Model

The linear spline analysis in Section 2.3.1 assumes that the observations are independent; however, independence is clearly not the case in repeated measures data. In these analyses, the correlation among response times within a subject needs to be taken into account. Linear mixed models are one tool that accounts for the correlation with a trial-constant, subject-specific response time effect. Please see the Phase I report for details about the specification of the linear spline mixed model.

For the separate models of sea level and above sea level subjects, the estimates are very similar to those in Table 2.2 which do not incorporate the subject random effect. The main difference is that the standard errors on the trial number and spline coefficients are smaller because some of the variability is captured by allowing subject-level variability via the random effects. As a result, two additional changes in slope (i.e. for knot coefficients at 25 and 55) are statistically significant for the above sea level subjects. The conclusions for the sea level trial profiles remains unchanged (i.e. the statistical significance remains the same but at a stricter level of confidence).

Table 2.3: Linear Spline Mixed Model Results

Covariate	Above Sea Level Estimate (s.e.)	Sea Level Estimate (s.e.)	Combined Estimate (s.e.)
Intercept	394.9 (28.02)**	344.6 (16.44)**	394.89 (23.08)**
Sea Level			-50.34 (33.72)
Trial Number (Slope: Trials 1-5)	-15.20 (6.45)**	-9.19 (3.63)**	-15.20 (5.27)**
Sea Level* Trial Number^			6.01 (7.66)
Spline: Knot at 5 (Change in Slope: 6-15)	19.03 (7.70)**	8.20 (4.35)*	19.02 (6.30)**
Spline: Knot at 15 (Change in Slope: 16-25)	-6.94 (3.28)**	1.95 (1.87)	-6.94 (2.68)**
Spline: Knot at 25 (Change in Slope: 26-35)	5.59 (3.08)*	-0.59 (1.77)	5.59 (2.52)**
Spline: Knot at 35 (Change in Slope: 36-45)	1.99 (3.19)	0.46 (1.85)	1.99 (2.61)
Spline: Knot at 45 (Change in Slope: 46-50)	-13.76 (5.54)**	0.89 (3.16)	-13.76 (4.53)**
Spline: Knot at 50 (Change in Slope: 51-55)	16.76 (7.73)**	-1.22 (4.38)	16.76 (6.32)**
Spline: Knot at 55 (Change in Slope: 56-65)	-9.65 (5.56)*	-1.90 (3.17)	-9.66 (4.55)**
Spline: Knot at 65 (Change in Slope: 66-75)	3.99 (3.35)	2.31 (1.94)	3.99 (2.74)
Spline: Knot at 75 (Change in Slope: 76-80)	-6.37 (6.18)	-0.75 (3.53)	-6.36 (5.06)
Sea Level * Spline: Knot at 5^			-10.82 (9.15)
Sea Level * Spline: Knot at 15^			**8.90 (3.92)**
Sea Level * Spline: Knot at 25^			**-6.18 (3.70)***
Sea Level * Spline: Knot at 35^			-1.53 (3.84)
Sea Level * Spline: Knot at 45^			**14.65 (6.63)**
Sea Level * Spline: Knot at 50^			**-17.99 (9.21)***
Sea Level * Spline: Knot at 55^			7.76 (6.65)
Sea Level * Spline: Knot at 65^			-1.69 (4.04)
Sea Level * Spline: Knot at 75^			5.60 (7.40)
Random Intercept Variance	2934	1108	2125

Residual Variance	12932	3700	8648

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

^ The additional change in slope for sea level subjects.

We can formally test for differences in mean profiles by fitting a model that has interaction terms for each altitude level. The results from this combined model are in Table 2.3. In this model, the change in slope for trials 16-25, 26-35, 46-50 and 51-54 are statistically significantly different for sea level versus above sea level subjects with estimates of 8.90, -6.18, 14.65, and -17.99, respectively. That is, the model picks up that the average trial profile for above sea level subjects has a different shape in these trial number ranges than the sea level subjects.

2.4   Conclusions

This section presents exploratory analysis and modeling results of sea level and above sea level subjects in the altitude data for SRT1 and SRT2 together. The linear spline mixed model of average trial profiles developed in Phase I of this series of reports is applied to the altitude data, with the addition of knot points for the additional set of 40 trials. We observe that a linear spline analysis captures the shape of the average trial profiles and indicates statistically significant differences in the shape of the average profiles for sea level and above sea level subjects. The results for the first set of 40 trials are largely consistent with findings from Phase I, with the exception of the tail end of the SRT1 trials where the model is affected by the larger average response times for SRT2. With the addition of the second set of 40 trials, the above sea level subjects exhibit a second learning effect depicted as a sharp downward slope in response time. Through the 80 trials, the above sea level subjects display a variable and wiggly average response time trial profile; while the sea level subjects show a relatively constant (after the initial learning effect) linear average response time trial profile.

3   Code Sub-Learning Assessments

In addition to SRT, DANA includes other tests. One of these is code sub-learning (CSL). An appealing aspect of CSL is that, by definition, it is a learning test in which performance is expected to be more favorable among individuals with greater learning capacity. People with diminished capacity (e.g. those with dementia, head injury, hypoxia, etc.) would be expected to perform poorly on this test relative to controls because of their diminished capacity to learn and perform a new task. We will extend the model-based strategy that was identified with SRT data to CSL data to determine if (1) differences in the shape of repeated trials can be observed for normal vs. stressed patients and (2) if these differences are more pronounced than there were for SRT. Once again, we will rely on the Altitude data set for this task, using each subject as their own control.

3.1 Data

This section presents repeated measures analysis of the code sub-learning (CSL) data from the Excel file "Altitude (CodeSub).xlsx" received by Resnick, Chodorow and Associates on September 18, 2015. This analysis makes some simplifying data assumptions:

- Only trials at 5260m above sea level ("altitude = 1") and at sea level ("altitude = 3") are included in the analysis.
- Only trials from "administration = 1" are included in the analysis – this corresponds to the CSL component of DANA.
- Subject 19 was administered the test twice above sea level ("altitude = 1"). Both sets of trials are included in this analysis, but are treated as independent administrations (i.e. they are assigned different IDs).
- Lapsed trials ("response = Lapse") are excluded from the analysis.

After these exclusions we are left with a dataset of 2,736 CSL trials uniquely identified by ID, trial number, and altitude. We focus our attention on these three variables in addition to reaction time, the primary variable of interest. Table 1 presents summary information about the analysis data set.

Table 3.1: Summary Information about the Altitude (CSL) Data

	Above Sea Level	Sea Level
Number of Subjects	21	17
Average Number of Trials Per Person	71.4	71.9
Range of Number of Trials Per Person	66-72	71-72
Average Response Time	1231.0	1154.1
Range of Response Time	557-2991	494-2924
Average of Subject-Average Response Time	1234.4	1154.2
Range of Subject-Average Response Time	886.8-1701.8	874.0-1555.9

3.2 Exploratory Data Analysis

Given the similarity in the nature of the CSL and SRT data, the exploratory data analysis of the trial profiles of response time for the CSL data will mirror that from the Phase I report. Just as in the Phase I report, the following exploratory graphs and statistical methods were used to select an appropriate mean structure for a linear model of the trial profiles.

Figure 3.1 plots the trial profiles of response time for each subject by altitude. Each line represents one subject's response times traced over up to 72 trials. The trial profiles appear erratic in that within altitude, there is no obvious visual trend that is common across subjects. Furthermore, it is even hard to discern an average trend at either altitude.
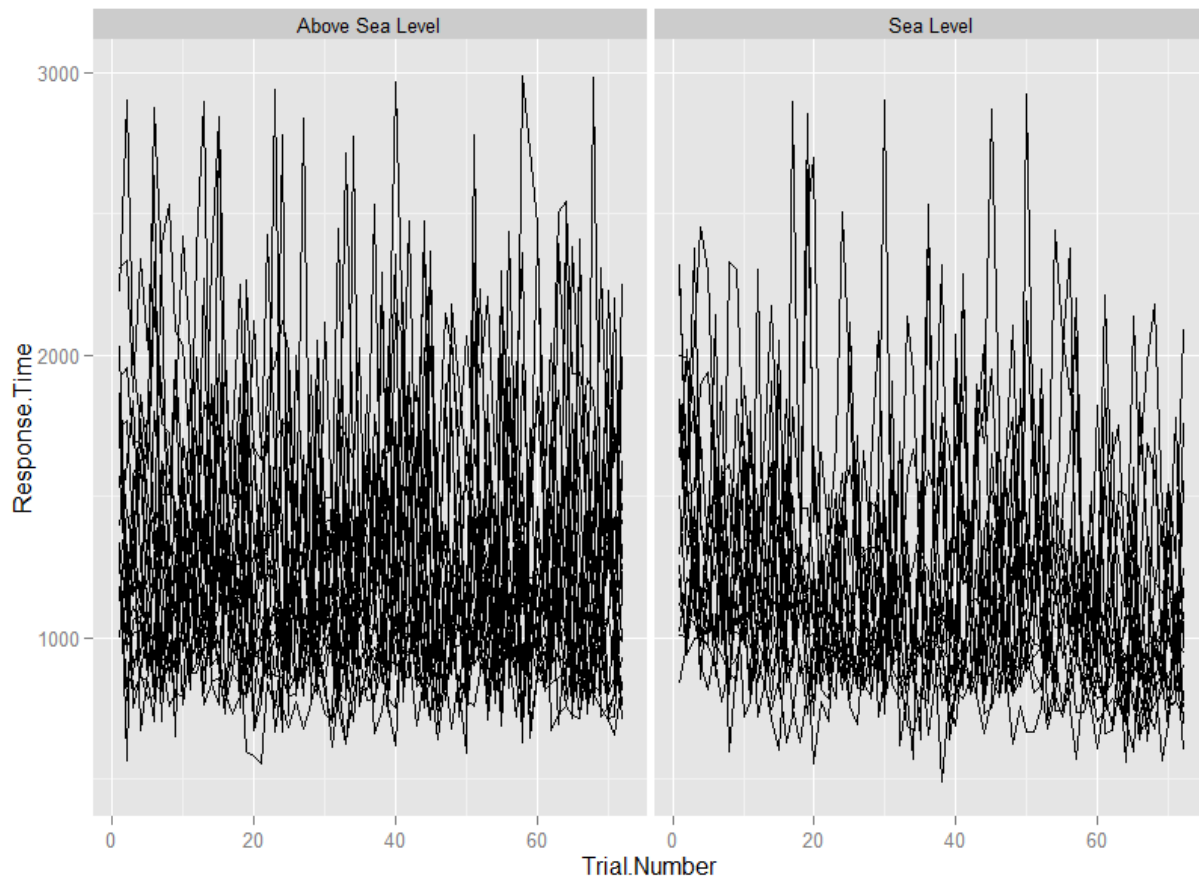
Figure 3.1: Trial Profiles of Response Time by Altitude

Figure 3.2 adds average response times by trial number (red triangles) to the trial profiles shown in Figure 3.1. The red triangles depict some interesting average trends. For example, a "learning effect" is visually apparent for both altitudes. The average response time for the initial trial is the largest observed average response time, followed by a decrease in average response time for both altitudes. With the exception of a cluster of lower average response times in the final trials for the sea level subjects, the red triangles do not illustrate any striking differences in the average trial profiles between the two altitudes.
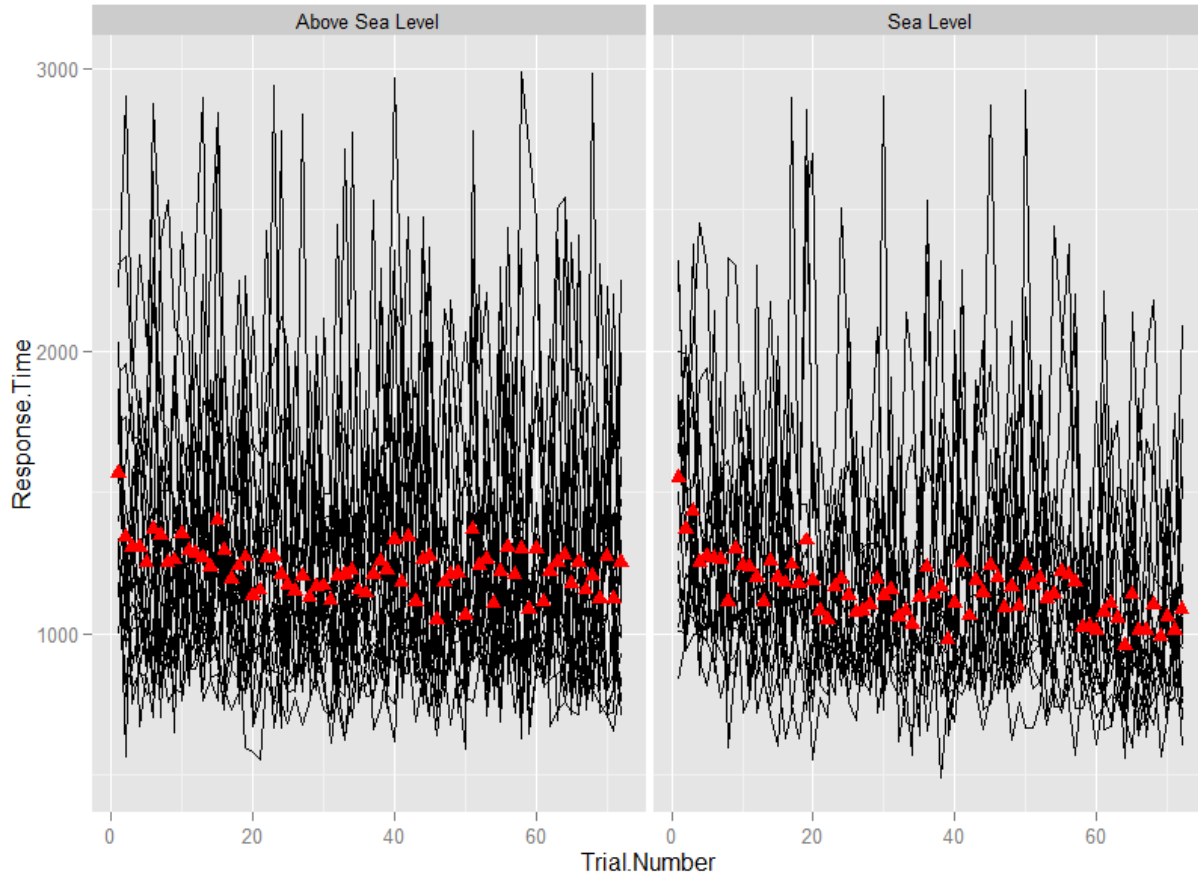
Figure 3.2: Trial Profiles by Altitude: Trial Averages

Figure 3.3 includes a loess curve fit to the trial data. This smooth fit provides an appealing graphical summary of the relationship between response time and trial number. While there are certainly humps in the curves, both curves exhibit a smooth overall shape. That is, it appears that wiggliness is often cause by small changes within small windows of trials, rather than representative of a larger trend. Imposing loess curves on the plots of trial profiles facilitates visualization of differences and – in the case of this CSL data – similarities. For both the above sea level subjects and the sea level subjects we observe: (1) similar variability around the loess curve, (2) a steep decrease in average response time in the initial trials, (3) a downward sloping trend for the first ~20 trials, and (4) a slight upward sloping trend in the mid-range of the trial number. The most apparent differences between the average trial profiles between sea level and above sea level subjects are in the last ~ 20 trials where there is a dip in average response time for sea level subjects, while the above sea level subjects show a stable linear trend.
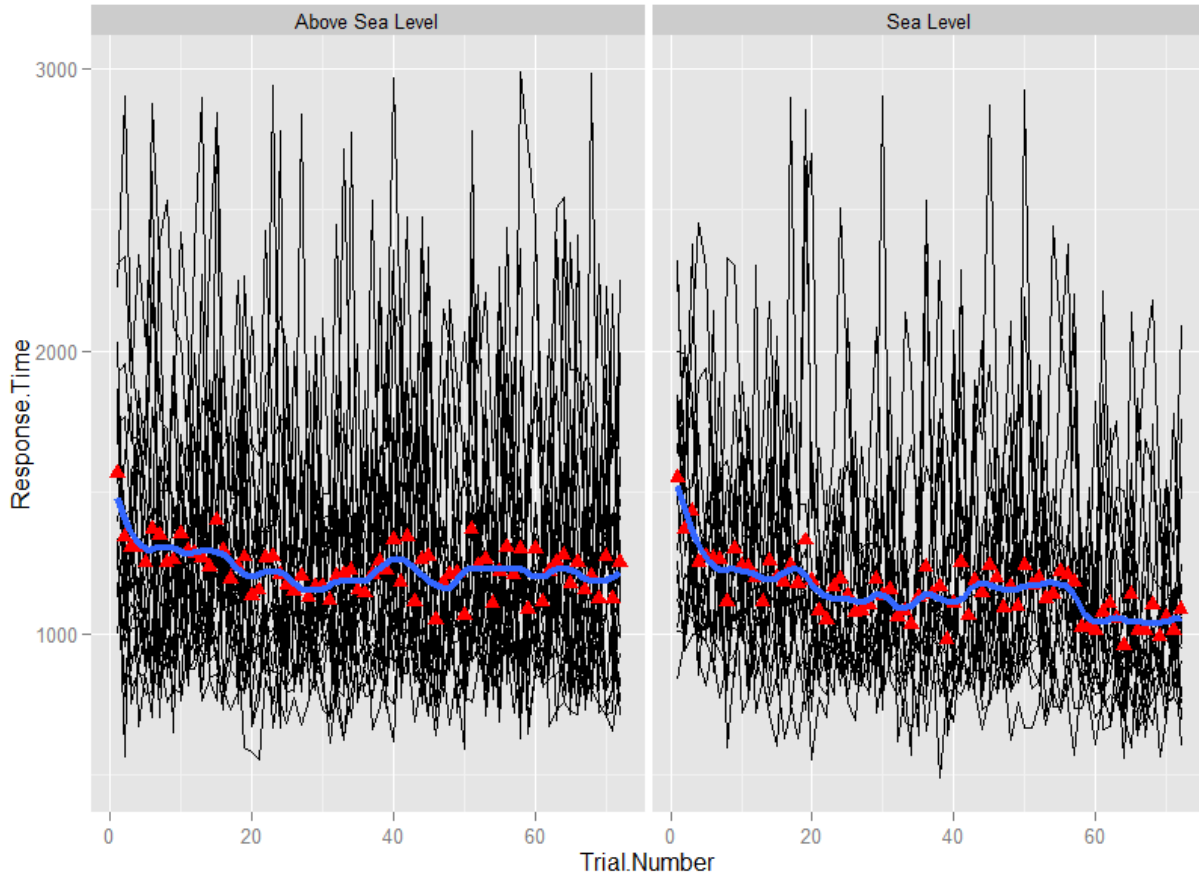
Figure 3.3: Trial Profiles by Altitude: Loess

3.3 Linear Models

Using observations from the exploratory data analysis and insight concerning the shape of the curve from the loess, we fit linear statistical models to capture the dependence between response time and trial number. Models of the two altitudes were assessed separately with the goal of identifying a generally applicable model.

3.3.1 Quadratic Regression

Looking past the small humps in the loess curves, the trends appear roughly linear with some curvature. A simple model to capture curvature is quadratic regression: a simple linear regression that includes a quadratic term for trial number.

We observe that this model captures a decent amount of overall curvature in above sea level subjects, but for the sea level subjects it misses a distinct change in slope between trials 40 and 60. For both altitudes, the quadratic model does not adequately fit the steep decline in average response time in the first few trials. Figure 3.4 shows that quadratic polynomial linear regression is useful for assessing the overall shape of the curves, but it misses some subtleties that may represent important differences.

14

Nonetheless, the smoothness of the green quadratic curves facilitates comparisons of trial profiles: downward sloping for both above sea level and sea level subjects, with the above sea level subject's trial profile leveling out while the sea level subject's trial profile continues to decrease.
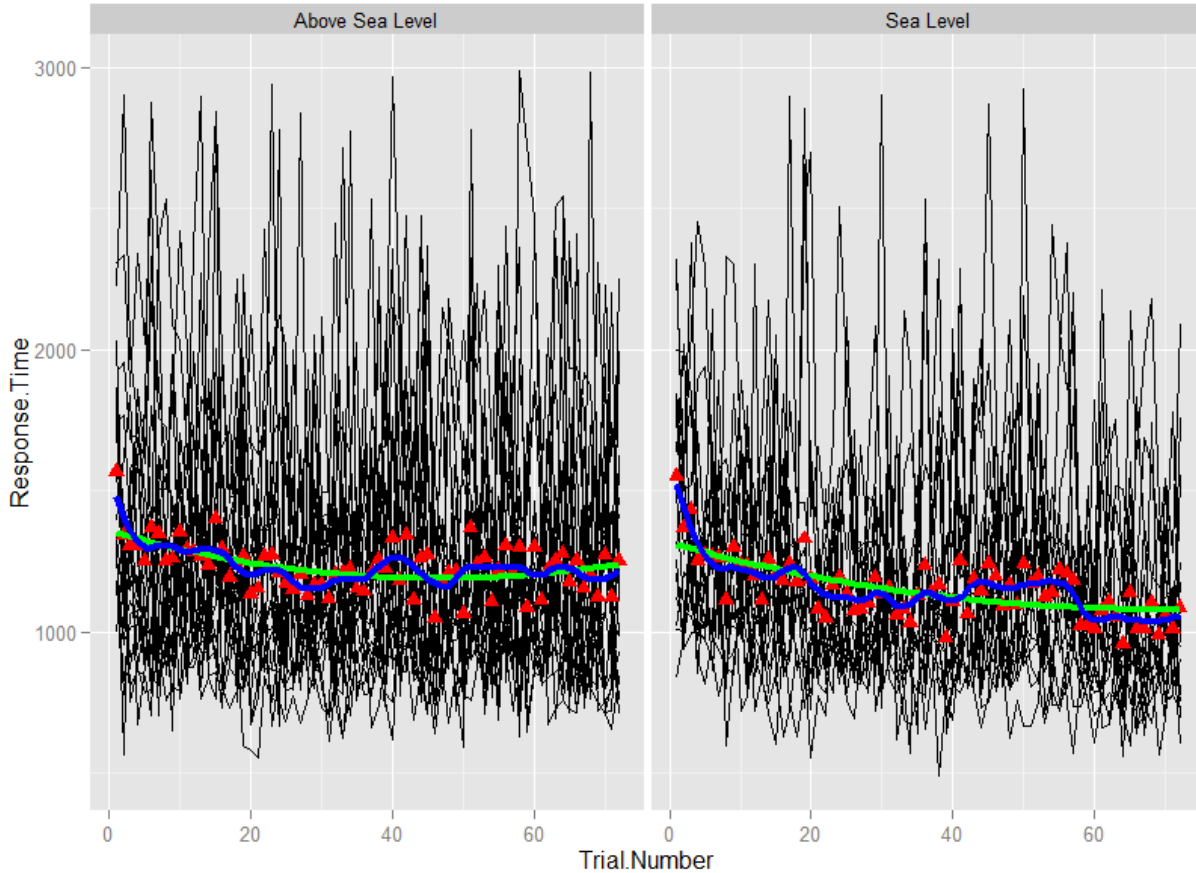

Figure 3.4: Trial Profiles by Altitude: Linear Regression with Quadratic (Green) vs. Loess (Blue)

3.3.2   Linear Spline Regression

Using the loess curves as a visual guide, the shape suggests that changes in slope occur after the first few trials, and every 10 or so trials in the mid-range of the number of trials. Accordingly, we chose knots at trial numbers 5, 30, 40 and 50. For details of linear spline regression, please see the Phase I report. Figure 3.5 shows that the linear spline regression with just four knot points captures the important features of the loess curves in Figure 3.3. That is, a set of just five piecewise linear regressions, reasonably depict the shape of the average trial profiles for above sea level and sea level subjects. In contrast to the SRT data, the differences in the CSL trends by altitude appear more in the steepness of slopes rather than the changes in the slope – as the wiggliness of the curves does not seem to tell a story.
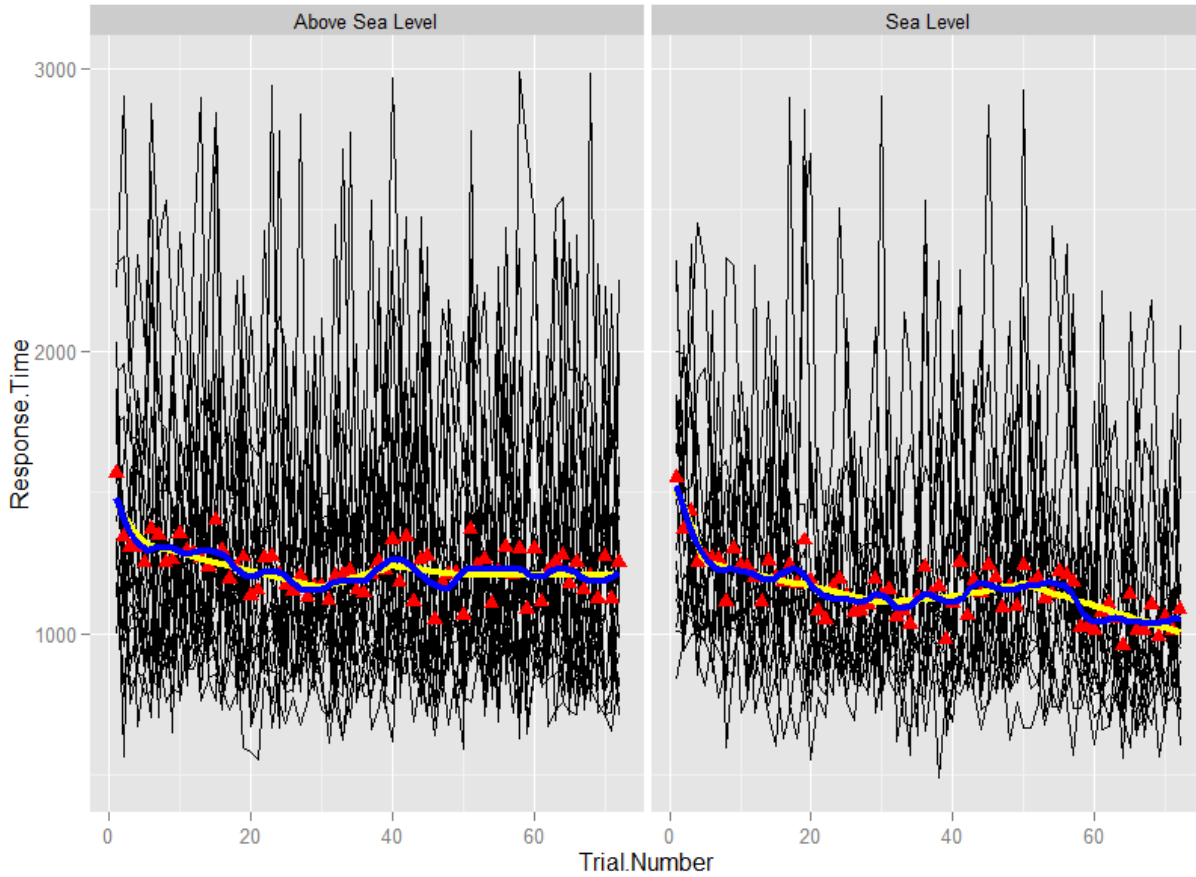
Figure 3.5: Trial Profiles by Altitude: Linear Regression with Splines (Yellow) vs. Loess (Blue)

Table 3.2 presents estimated coefficients, standard errors, and statistical significance from the linear spline fit with a subject random intercept. This linear mixed model accounts for the correlation within a subject with a trial-constant subject-specific response time effect. In this analysis, we introduce only a random intercept into the linear model. This allows each subject's intercept (average response time) to be different from the others. Similar to previous reports, incorporating the random intercept has little effect on the coefficients, but reduces standard errors because the random effect accounts for some of the variability of the linear spline regression.

Table 3.2: Linear Spline Mixed Model Results

Covariate	Above Sea Level Estimate (s.e.)	Sea Level Estimate (s.e.)	Combined Estimate (s.e.)
Intercept	1486.9 (87.20)**	1570.9 (84.48)**	1486.92 (82.38)**
Sea Level			84.01 (123.16)
Trial Number (Slope: Trials 1-5)	-33.12 (17.11)*	-63.30 (16.62)**	-33.12 (16.18)**
Sea Level * Trial Number^			-30.18 (24.19)
Spline: Knot at 5 (Change in Slope: 6-30)	26.76 (18.02)	57.52 (17.50)**	26.76 (17.04)
Spline: Knot at 30 (Change in Slope: 31-40)	14.83 (5.75)**	7.94 (5.56)	14.83 (5.44)**
Spline: Knot at 40 (Change in Slope: 41-50)	-12.18 (8.46)	2.45 (8.07)	-12.18 (7.91)

16

Spline: Knot at 50 (Change in Slope: 51-72)	3.75 (6.19)	-12.40 (5.99)**	3.75 (5.85)
Sea Level * Spline: Knot at 5^			30.76 (25.47)
Sea Level * Spline: Knot at 30^			-6.89 (8.11)
Sea Level * Spline: Knot at 40^			14.63 (11.77)
Sea Level * Spline: Knot at 50^			**-16.15 (8.73)***
Random Intercept Variance	44271	33125	39318
Residual Variance	129724	99169	116005

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

^ The additional change in slope for sea level subjects.

At both altitudes, there is a statistically significant negative slope between trials 1 and 5: -33.12 and -63.30 for above sea level and sea level subjects, respectively. For the sea level model, there is a statistically significant change in slope at trial 5 and trial 50. Both of these changes in slope result in an average trial profile that continues to be negatively sloped, but less so as compared to the initial "learning effect." For example, the initial slope of -63.30 changes by 57.52 to -5.78 between trials 6 and 30: -63.60 is much more negative than -5.78. For the above sea level model, the only statistically significant change in slope is at trial 30, where the trial profile goes from having a negative slope (-6.36) to a positive slope (8.47); this is about the point where the trial profile levels out.

These results suggest possible differences in average response time patterns for subjects at different altitudes. Both altitudes exhibit an initial learning effect; the average response times show a downward slope in the first 5 or so trials. However, while the above sea level subjects level-out on average, the sea level subjects exhibit a continued learning, particularly in the last 20 trials.

We can formally test for differences in mean profiles for each altitude by fitting one model that has interaction terms for altitude level (see Table 3.2, the "combined" column, for estimated coefficients, standard errors and statistical significance). In this model, only the change in slope at the last knot point is statistically significantly different between altitudes. At trial 50, the sea level subjects exhibit a large negative change in slope, while the above sea level subjects show a small positive change (which is not statistically significant). The difference in the initial "learning effect" between about sea level and sea level is not statistically significant, even though it is quite large (30.76). Figure 3.6 presents the estimated linear spline fit for above sea level and sea level subjects in the CSL altitude data. The difference in scale, as compared to Figure 3.5, clarifies the similarities and differences in levels and changes of slope.
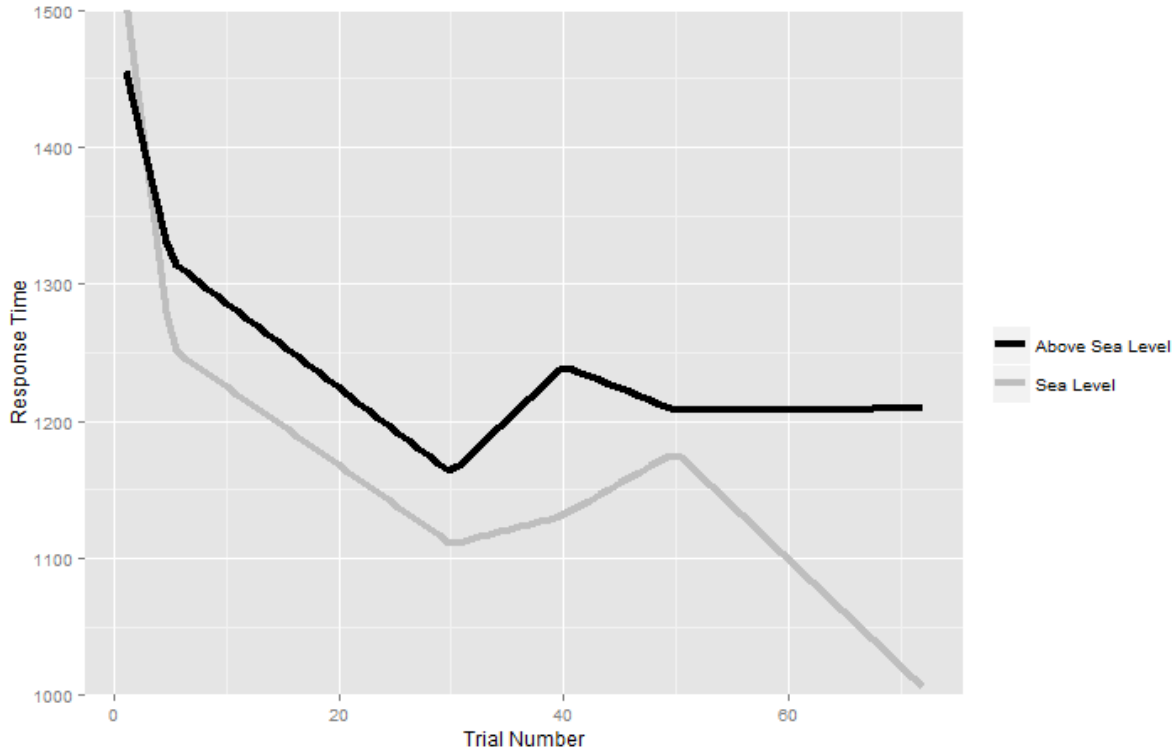
Figure 3.6: Linear Spline Fit of Response Time by Altitude

In contrast to the results from analysis of the SRT data, the interesting differences in response time trial profiles for the CSL data are in the differences in the steepness of the slopes rather than the shape of the curves (i.e. the curvature and the wiggliness). Table 3.3 presents the value and statistical significance of the slope for each set of trials (as defined by the knot points). The results are consistent with the results and discussion of changes in slope from the linear spline mixed model. That is, the only statistically significant difference in slope between sea level and above sea level subjects is after trial 50, where the slope for above sea level subjects is .03 and for sea level subjects is -7.79 (which is statistically significantly different than zero). As before, while the slope of the initial "learning effect" is sharply negative and statistically significant for each altitude, the difference in this slope between altitudes is large but not statistically significant.

Table 3.3: Test of Slopes from Linear Mixed Model

	Above Sea Level	Sea Level	Combined
Trial Number Range	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
1-5	-33.12 (17.11)*	-63.30 (16.62)**	-30.18 (24.19)
6-30	-6.37 (1.80)**	-5.78 (1.74)**	0.59 (2.54)
31-40	8.46 (4.50)*	2.16 (4.35)	-6.30 (6.34)
41-50	-3.72 (4.61)	4.61 (4.46)	8.34 (6.50)
>50	0.03 (2.21)	-7.79 (2.14)**	-7.82 (3.11)**

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively

Looking at Figure 3.6 and Table 3.3 together, we see that the average trial profiles are more or less parallel, with the sea level subjects exhibiting a lower average response time throughout the trial period.

18

The most striking difference is the divergence in the average trial profiles at the end of the trial period, where above sea level subjects maintain a flat profile (a not statistically significant slope of 0.03) while average response time for sea level subjects significantly declines (a statistically significant slope of -7.79).

3.4   Comparison to SRT Linear Spline Model

While the nature of the SRT and CSL data are similar, they are administered to measure different characteristics of a subject's cognitive capacity. As such, via exploratory data analysis, we assessed the characteristics of the CSL response time trial profiles independently of model fits from previous analyses of the SRT data. While the exploratory data analysis of the CSL data also indicated localized curvature leading to linear spline models, the chosen knot points differ in the analysis of the CSL and SRT data. The average response time trends appear less variable with fewer localized slope changes in the CSL data. In fact, a linear spline fit that allowed just four changes in slope (i.e. knot points) captures the overall trend. For the SRT data with the full set of 80 trials, the linear spline model allows eight changes of slope. While there are differences in the placement of these knot points, it is important to note that both the SRT and CSL models have a knot point to capture the initial "learning effect" (a knot for trial number 5), and a set of points in the mid-range of trial number. Figure 3.7 shows the linear spline fit from applying the linear spline model used for the SRT data to the CSL data. Certainly this model with more knot points fits the curves well, but is somewhat over-fit as the additional knots are not necessary to tell the same story.
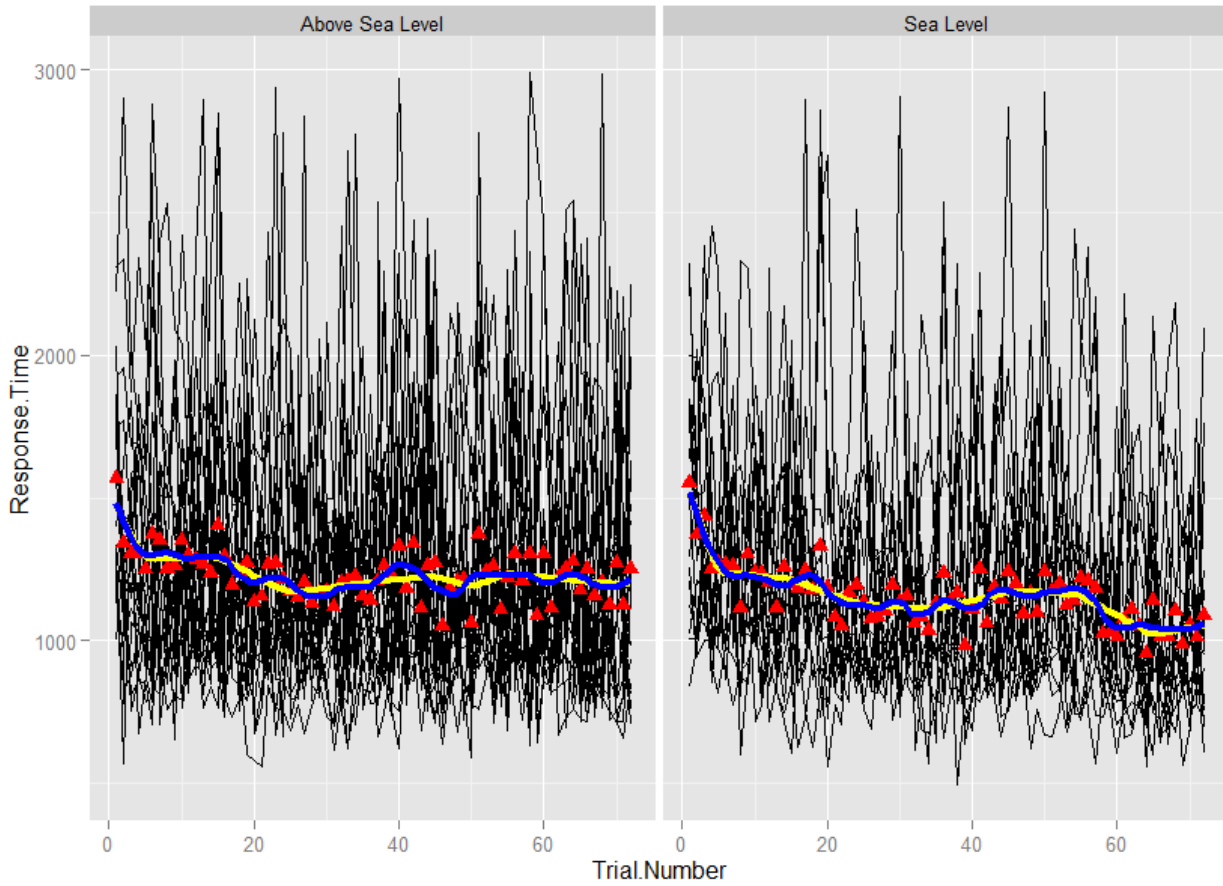
Figure 3.7: Trial Profiles by Altitude: Linear Regression with SRT Splines (Yellow) vs. Loess (Blue)

For comparison purposes, the results from the SRT linear mixed model applied to the CSL data are presented in Table 3.4. We observe few statistically significant changes in slope. This is consistent with the previous findings that the shape and "wiggliness" of the curves in the CSL data are of less interest than the differences in the steepness of the slope. This indicates that not only does the use of too many knot points over-fit the curve, but this approach also masks the differences in the slope of the curve in the last 20 trials for the sea level and above sea level subjects. Thus, while the statistical methodology used to analyze the SRT data are applicable to the CSL data, it is important to assess the average response time trial profiles independently to best capture differences in the trends.

Table 3.4: Linear Spline Mixed Model Results

Covariate	Above Sea Level Estimate (s.e.)	Sea Level Estimate (s.e.)	Combined Estimate (s.e.)
Intercept	1512.34 (90.24)**	1579.82 (86.35)**	1512.34 (85.02)**
Sea Level			67.47 (126.65)
Trial Number (Slope: Trials 1-5)	-44.12 (19.51)**	-67.75 (18.84)**	-44.12 (18.41)**
Sea Level * Trial Number^			-23.64 (27.52)
Spline: Knot at 5 (Change in Slope: 6-15)	44.73 (23.36)*	65.03 (22.55)**	44.73 (22.04)**

Spline: Knot at 15 (Change in Slope: 16-25)	-13.23 (10.06)	-6.30 (9.68)	-13.23 (9.49)
Spline: Knot at 25 (Change in Slope: 26-35)	16.06 (9.52)*	7.54 (9.16)	16.06 (8.98)*
Spline: Knot at 35 (Change in Slope: 36-45)	-0.53 (9.90)	7.37 (8.53)	-0.53 (9.34)
Spline: Knot at 45 (Change in Slope: 46-50)	-11.94 (17.06)	-5.73 (16.35)	-11.94 (16.09)
Spline: Knot at 50 (Change in Slope: 51-55)	19.48 (23.62)	-2.50 (22.71)	19.48 (22.29)
Spline: Knot at 55 (Change in Slope: 56-65)	-13.50 (17.34)	-11.37 (16.67)	-13.50 (16.36)
Spline: Knot at 65 (Change in Slope: 66-80)	0.75 (14.45)	18.77 (13.88)	0.75 (13.63)
Sea Level * Spline: Knot at 5$^\wedge$			20.30 (32.95)
Sea Level * Spline: Knot at 15$^\wedge$			6.93 (14.16)
Sea Level * Spline: Knot at 25$^\wedge$			-8.51 (13.40)
Sea Level * Spline: Knot at 35$^\wedge$			7.90 (13.74)
Sea Level * Spline: Knot at 45$^\wedge$			6.21 (23.96)
Sea Level * Spline: Knot at 50$^\wedge$			-21.99 (33.24)
Sea Level * Spline: Knot at 55$^\wedge$			2.13 (24.40)
Sea Level * Spline: Knot at 65$^\wedge$			18.01 (24.40)
Random Intercept Variance	44682	33122	39395
Residual Variance	131516	99296	117062

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.
$^\wedge$ The additional change in slope for sea level subjects.

3.5   Conclusion

This section presents exploratory analysis and modeling results of sea level and above sea level subjects in the CSL altitude data. We observe that a linear spline analysis – with just a few points of slope change – captures the overall shape of the trial profiles. There is some wiggliness in the shape of the curves; however, the more distinct differences between sea level and above sea level subjects appear in the magnitudes of the slopes, rather than the changes in slope. The "learning effect" is observed at both altitudes, but is steeper for sea level subjects (though the difference is not statistically significant). Furthermore, the above sea level subjects show a relatively steady decrease then plateau of average response time, while the sea level subjects exhibit continued "learning" in the last 20 trials. These differences are harder to distinguish via the estimated model coefficients when simply applying the SRT model to the CSL data, indicating that a tailored set of knot points in a linear spline regression should be considered for different measures of cognitive ability.

Simple Reaction Time Repeated Measures Analysis: Phase IV Report

This memo summarizes recent results from ongoing work on the "Repeated Measures" project. The overarching goal of this project is to identify quantitative methods that can be used to distinguish a "normal" from a "non-normal" individual based on DANA response patterns. These methods could ultimately be used to identify individuals whose cognitive efficiency patterns have been unfavorably impacted by age-related cognitive decline, sleep disturbances, depression, sports-related head injury, or battle-related head injuries.

SUMMARY OF THE PHASE IV REPORT

- The overall objective of the work is to identify a group or groups of subjects whose repeated testing results differ from a group of "normal" subjects.
- Four data sets were used in various ways.
  - Ft. Hood data set
    - "Healthy" (n= 219; CES <= 8, no head injury, PHQ <= 9, and PCL < 50.)
    - "Unhealthy," (n=98)
  - Air Force data set
    - "Normal" cadets (n=153)
    - Cadets reporting concussion (n=6)
  - Altitude data set
    - People at sea level (n=17)
    - People at extreme altitude (n=21)
  - Burke/aging data
    - Healthy seniors (n=22)

- Alzheimer's patients (n=10)
- Two statistical approaches were used.
  - k-means clustering
  - Group-based trajectory modeling
- For each statistical approach, we first forced the "normal" data into two groups, and then we forced the same data into three groups.
  - N=389 (normals from Ft. Hood, Air Force, and altitude)
  - N=170 (normals from Air Force and altitude)
- After identifying the clusters using "normal" data, we brought in data for "non-normals" (e.g. "unhealthy" Ft. Hood service members, concussed cadets, hypoxic subjects, Alzheimer's patients).
  - Goal: Determine how many "non-normal" people from each data set fall into each previously-defined cluster that was based on "normal" data.
  - How many "known non-normal" people will be classified into the worst "normal" group?
- Most work was conducted with SRT data; some was conducted with CSL data.

BOTTOM LINE AND POTENTIAL NEXT STEPS
- Results from the two statistical methods were very similar.
  - Both methods identify large groups with lower and stable mean response times as well as a small group with mean response times that are both longer and more variable over time.
  - The fact that the two methods largely agree on the clusters that are hidden in the data indicates that they (the clusters) are relatively solid in a statistical sense.
  - It is unlikely that we need to expand the pool of normals to refine the clusters because our results were very similar when we examined clusters that were based on 170 normals and clusters that were based on 389 normals.
- These results are preliminary because the models have not been validated.
- These results are heavily focused on SRT.
  - These methods can be easily applied to other DANA tests.
- A classification rule can be developed using the existing data and will require the following steps:
  - Divide the data into a "development or testing" data set and a "validation" data set.
    - A validation group typically consists of 25% of subjects in the full data set, and the development/testing group is the other 75%.
  - Develop a model to identify "non-normal" subjects using the 75% of subjects that comprise the "development" group.
    - Include covariates (e.g. age, medial history) that are predictive of group membership

- - Apply this model to the 25% of subjects in the validation data set and classify each subject into the appropriate group.
    - Compute measures of fit for this group.

ISSUES FOR CONSIDERATION FOR NEXT STEPS

- As a group, we decide how sensitive and specific any future decision rule will be based on how we approach these next steps.
- In developing a decision rule, do we aim for "larger" or "smaller" groups of "non-normals"?
  - Sensitivity vs. specificity?
  - What percentage of "non-normals" is realistic and/or clinically appropriate?
    - Does this percentage differ in various settings?
  - How do we "value" false positives vs. false negatives?
    - How might future users value this tradeoff?
- There are a number of ways that DANA tests can be used individually and in combination for development of a decision rule that ultimately classifies an individual as normal or not normal.
  - We could set up a two-stage screening process where we select an initial test(s) to potentially identify large numbers of subjects for further screening, and do a second test to create a "tighter" group of "non-normals."
    - Use the first test to cast a "wide net"
    - Use additional tests to reduce the size of the net (enhance specificity)
  - We could select several DANA tests and use these tests to develop a summary score that is used for classification.
    - We could force the data for each selected test into three clusters
    - Assign a score of 0, 1, or 2 to each individual for each test
    - Sum an individual's scores and develop a decision rule based on the summary score
  - Other options can also be explored or developed
- How many DANA subtests do we want to use for the purpose of identifying non-normals?
  - This has potential implications for the complexity of the rule
  - A more complex rule may only perform marginally better than a simpler one

OVERVIEW

The goal of these analyses is to identify a group or groups of subjects whose repeated testing results differ from a group of "normal" subjects. In practice, it is often a challenge that subjects are not predefined as "normal" or "abnormal." This leads to the need to look for groups within a data set that are similar in behavior over the course of a given longitudinal trajectory. We selected strategies that rely on statistical methods that fall in the category of unsupervised learning techniques. This is a type of machine learning algorithm that is used to draw inferences from datasets consisting of input data without labeled responses (e.g. normal or non-normal). These approaches are used for exploratory data analysis to find hidden patterns or groupings in data. With this approach, the current state (e.g., normal, abnormal) of the individual is not used to identify predictors of that state; rather the data are divided in a systematic manner into groups that behave similarly.

The analyses in this report used two statistical methods: k-means clustering and group-based trajectory modeling. Results from each method are presented separately, and a comparison of the two is presented at the end of the report.

K-MEANS CLUSTERING

The goal of k-means clustering is to partition a population of n subjects into k groups whose trajectories are similar to each other. Each longitudinal trajectory is then placed into the cluster to which it is "closest." This method is widely used in data mining and genetic analysis. It is also straightforward to apply and allows for a mechanism to classify any new subjects who were not part of the original analysis sample. The latter feature is directly relevant to the ultimate goal of this line of investigation. To classify a new subject, the distance between the new subject's trajectory and the center of each cluster is calculated, and the subject is then classified into the group to which it is closest. A notable disadvantage of the k-means clustering approach is that potentially important predictors such as age or health history cannot be used as part of the process by which subjects are partitioned into groups.

K-means clustering was applied to both the SRT and CSL data sets for the Ft. Hood, Air Force, and altitude data sets, and it was implemented using the kml package in R. In these analyses, the data for "normal" people in the Ft. Hood, Air Force, and altitude data sets were forced into two groups, and then into three groups.

In addition to providing estimated clusters of subjects' longitudinal profiles, the clusters that result from k-means can be used to predict cluster membership of individuals who were not included in the estimation of the k-means clusters (i.e. "out-of-sample predictions"). For example, we can first identify clusters from the full set of 389 "normal" subjects, then use this information to predict cluster membership of the seniors with Alzheimer's who were in the Burke study.

This prediction is done by calculating the Minkowski distance between a subject's individual trajectory and the center of the cluster that is calculated for each group. In this report, we look at two cases of the Minkowski distance for determining out-of-sample prediction: Euclidean (p=2) and Manhattan (p=1). The Minkowski distance is defined as:

$$d(y_i, \overline{y_k}) = \left(\sum_{t=1}^{T} |y_{it} - \overline{y_{kt}}|^p \right)^{1/p}$$

where y_i is the vector of 40 simple reaction times for subject i and $\overline{y_k}$ is the vector of 40 mean simple reaction times calculated from the subjects classified in cluster k. These vectors have elements y_{it} and $\overline{y_{kt}}$, the SRT measurement at trial number t. Each subject is then classified into the group to which it is closest, i.e. where $d(y_i, \overline{y_k})$ is smallest.

This classification is a simple decision rule for identifying "non-normal" subjects – it finds those that look most like the "non-normal" group among the "normal" subjects that are identified via the k-means clustering procedure. These classifications provide insight as to how useful the information from the estimated clusters is in distinguishing "normal" from "non-normal" subjects.

SRT results using k-means clustering

Two sets of k-means clustering analyses were fit to two sets of "normal data". We initially had access to "normal" data from the Air Force and altitude data sets. Subsequently, we more than doubled the number of normals by adding normal from the Ft. Hood data set. Results are presented for both sets of analyses.

- The "full normal data set" of 389 subjects:
  - 219 "normal" Ft. Hood subjects
  - 153 non-concussed Air Force cadets
  - 17 subjects at sea level
- A "reduced normal data set" of 170 "normal: subjects that excluded the Ft. Hood data
  - 153 non-concussed Air Force cadets
  - 17 subjects at sea level

Full normal data set analysis (N = 389 subjects)

Two-group analysis

When the data are forced into two groups, the k-means cluster analysis assigns 75.6% of the 389 "normal" subjects to cluster A and the remaining 24.4% of the subjects to cluster B. Cluster A represents subjects with lower and stable mean response times, whereas cluster B captures subjects with higher and slightly more variable mean response times. The plot below shows the estimated mean trajectories for the two clusters that were obtained from the k-means clustering analysis.

kml: Healthy, Sea Level & Ft. Hood, 2 Clusters
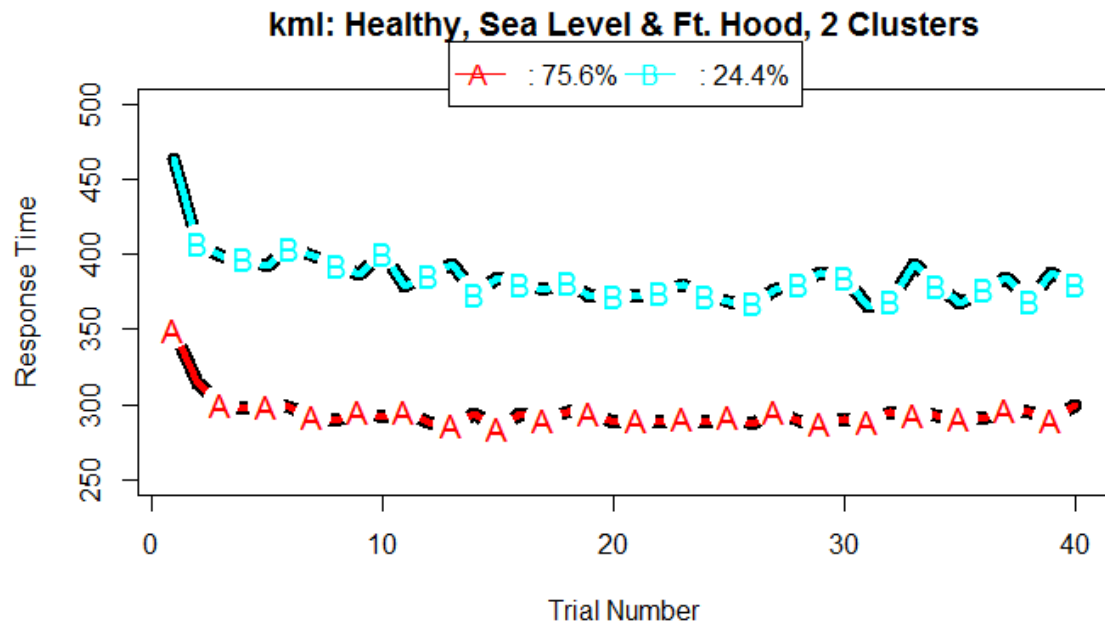
Using these two clusters, we then predicted the cluster assignment for the out-of-sample/"non-normal" subjects in various datasets. These individuals included the above sea level altitude subjects, concussed Air Force cadets, "non-normal" Ft. Hood subjects, healthy seniors, and seniors with Alzheimer's.

Table 1 presents data on how the out-of-sample/non-normal subjects were classified into cluster A vs. cluster B, using either the Manhattan or Euclidean distance. We observe that the Euclidean distance generally classifies a similar or larger proportion of the "non-normal" subjects to cluster B (the cluster with higher and more variable mean SRT response times) than the Manhattan distance.

Overall, all groups of "non-normal" subjects are classified into cluster B in higher numbers than the "normal" subjects on which the clustering analysis is based. That is, 24.4% of the subjects on which the clusters were defined were placed in cluster B (see figure above), but larger proportions of all the groups that are known to be non-normal fall into cluster B using the Euclidean distance. We find that all senior subjects are classified into cluster B, whereas between about 33% and 48% of subjects from the other datasets are classified in cluster B.

Table 1: Out-of-Sample Class Predictions, 2 Clusters, Full Data				
	Manhattan Distance		Euclidean Distance	
Group	A	B	A	B
Above Sea Level	16 (76.2%)	5 (23.8%)	11 (52.4%)	10 (47.6%)
Concussed Cadets	5 (83.3%)	1 (16.7%)	4 (66.7%)	2 (33.3%)
Ft. Hood Non-Healthy	74 (75.5%)	24 (24.5%)	61 (62.2%)	37 (37.8%)
Healthy Seniors	0 (0%)	22 (100%)	0 (0%)	22 (100%)
Alzheimer Seniors*	1 (10%)	9 (90%)	0 (0%)	10 (100%)

* Based on only 25 trials.

Three-group analysis

When the data are forced into three groups, the k-means cluster analysis assigns 53.7% of the 389 "normal" subjects to cluster A, 41.9% to cluster B, and 4.4% to cluster C. Cluster A contains subjects with lower and stable mean response times, cluster B captures subjects with higher but still relatively stable mean response times, and cluster C has subjects with larger and more variable mean response times. The plot below shows the estimated mean trajectories for the three clusters obtained from the k-means clustering analysis.
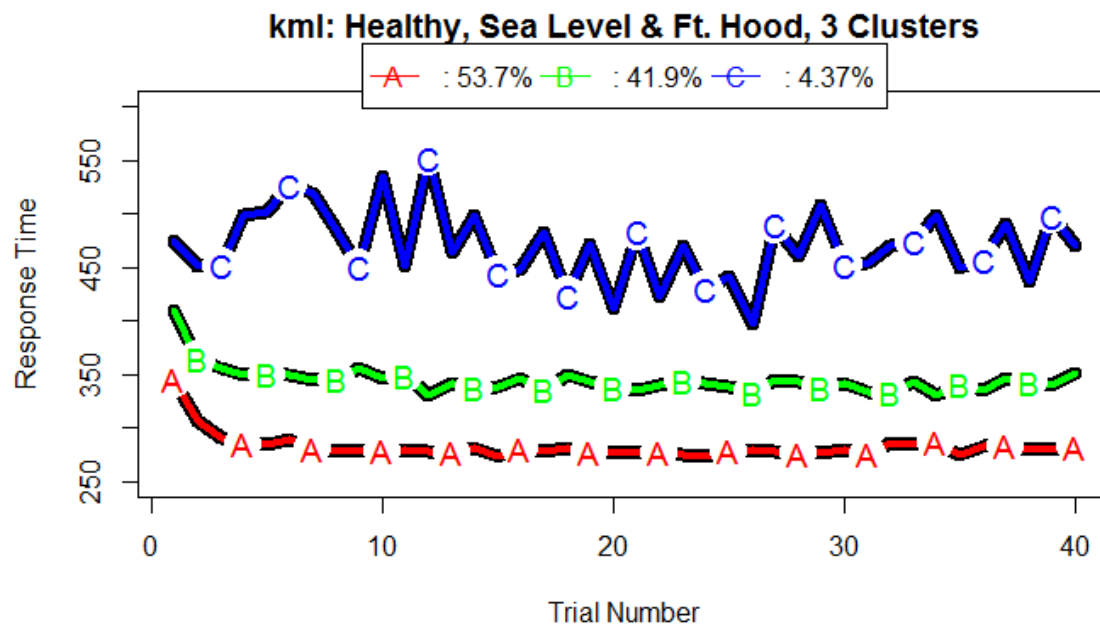
Once again, we look at out-of-sample subjects with a known, non-normal feature to see how they are classified across these groups.

Table 2 presents the number of each set of subjects classified into clusters A, B, and C assuming either the Manhattan or Euclidean distance. As in the two-cluster analyses, we observe that the Euclidean distance generally classifies a similar or smaller proportion of the "non-normal" subjects to cluster A (the most "normal" cluster) than the Manhattan distance. We see that all senior subjects are classified into either cluster B or C, where the split between cluster B and C is close to 50/50. On the other hand, between about 38% and 50% of subjects from the other datasets are classified in cluster A, where the majority of subjects classified into the "non-normal" clusters (B and C) are assigned to the intermediate cluster B.

Overall, as compared to the two-group analysis, the three-group analysis classified a smaller proportion of "non-normal" subjects into the most "normal" cluster (cluster A). That is, the three group approach placed more non-normal people in the less favorable performance clusters. Furthermore, these "non-normal" subjects are classified into cluster A at a lower rate than the "normal" subjects (53.7% in Figure above) on which the clustering analysis is based, although this difference for the concussed cadets and Ft. Hood non-healthy subjects is small.

Table 2: Out-of-Sample Class Predictions, 3 Clusters, Full Data						
	Manhattan Distance			Euclidean Distance		
Group	A	B	C	A	B	C
Above Sea Level	12 (57.1%)	7 (33.3%)	2 (9.5%)	8 (38.1%)	11 (52.4%)	2 (9.5%)
Concussed Cadets	3 (50.0%)	2 (33.3%)	1 (16.7%)	3 (50.0%)	2 (33.3%)	1 (16.7%)
Ft. Hood Non-Healthy	60 (61.2%)	36 (46.7%)	2 (2.0%)	49 (50.0%)	45 (45.9%)	4 (4.1%)
Healthy Seniors	0	9 (40.9%)	13 (59.1%)	0	9 (40.9%)	13 (59.1%)
Alzheimer Seniors*	0	5 (50.0%)	5 (50.0%)	0	5 (50.0%)	5 (50.0%)

* Based on only 25 trials.

Reduced Data Set: Air Force and Altitude data analysis (N = 170 subjects)

Two-group analysis

When two groups are forced, the k-means cluster analysis assigns 83.5% of the 170 subjects to cluster A and 16.5% of the subjects to cluster B. A slightly larger proportion of subjects are assigned to cluster A in this reduced set analysis than in the earlier analysis that included the Ft. Hood normal in the full data set (83.5% vs. 75.6%). Similar to findings from analyses of the full data set, Cluster A contains subjects with lower and stable mean response times, whereas cluster B captures subjects with higher and slightly more variable mean response times. The plot below shows the estimated mean trajectories for the two clusters obtained from the k-means clustering analysis of the reduced (n=170) data set.

kml: Healthy and Sea Level, 2 Clusters

A : 83.5%   B : 16.5%
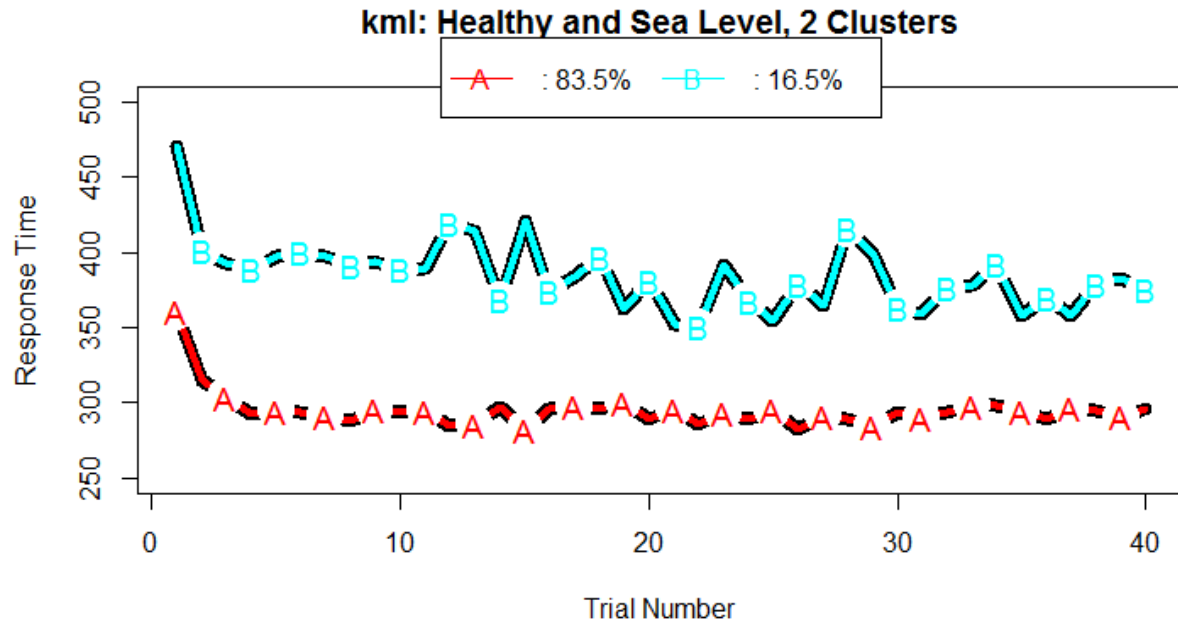
We then determined how frequently the clusters that were derived from the reduced data set place non-normals in cluster B. Table 3 presents the number of each of the non-normal groups that was classified into cluster A vs. cluster B. The Euclidean distance results are very similar to those we observed from the full set of data (n=389), with the exception of a slight decrease in the cluster B assignments of "non-healthy" Ft. Hood subjects. This finding indicates that the decision rule performs slightly better for Ft. Hood non-normals when Ft. Hood "normal" subjects are included in the analyses that define the clusters.   Because the Ft. Hood data is not used in the estimation of the k-means clusters in the reduced data set, we can use these data to predict the cluster assignment of the "healthy" Ft. Hood subjects. We find that over 70% of these "normal" subjects are classified into cluster A. Interestingly, the remainder of the Ft. Hood "normal" group is classified in cluster B.

Table 3: Out-of-Sample Class Predictions, 2 Clusters, Reduced Data				
	Manhattan Distance		Euclidean Distance	
Group	A	B	A	B
Above Sea Level	16 (76.2%)	5 (23.8%)	11 (52.4%)	10 (47.6%)
Concussed Cadets	5 (83.3%)	1 (16.7%)	4 (66.7%)	2 (33.3%)
Ft. Hood Healthy	170 (76.9%)	51 (23.1%)	158 (71.5%)	63 (28.5%)
Ft. Hood Non-Healthy	75 (76.5%)	23 (23.5%)	65 (66.3%)	33 (33.7%)
Healthy Seniors	0 (0%)	22 (100%)	0 (0%)	22 (100%)
Alzheimer Seniors*	1 (10%)	9 (90%)	0 (0%)	10 (100%)

* Based on only 25 trials.

<u>Three-group analysis</u>

When the reduced data set is forced into three groups, the k-means cluster analysis assigns 56.5% of the 170 subjects to cluster A, 40.6% to cluster B, and 2.94% to cluster C. These assignments are similar to the proportions observed from the full (n=389) set analysis. As in the full data analysis, cluster A contains subjects with lower and stable mean response times, cluster B captures subjects with higher but still relatively stable mean response times, and cluster C contains subjects with even higher and very wiggly mean response times. The plot below shows estimated mean trajectories for the three clusters obtained from the k-means clustering analysis for the n=170 data set.
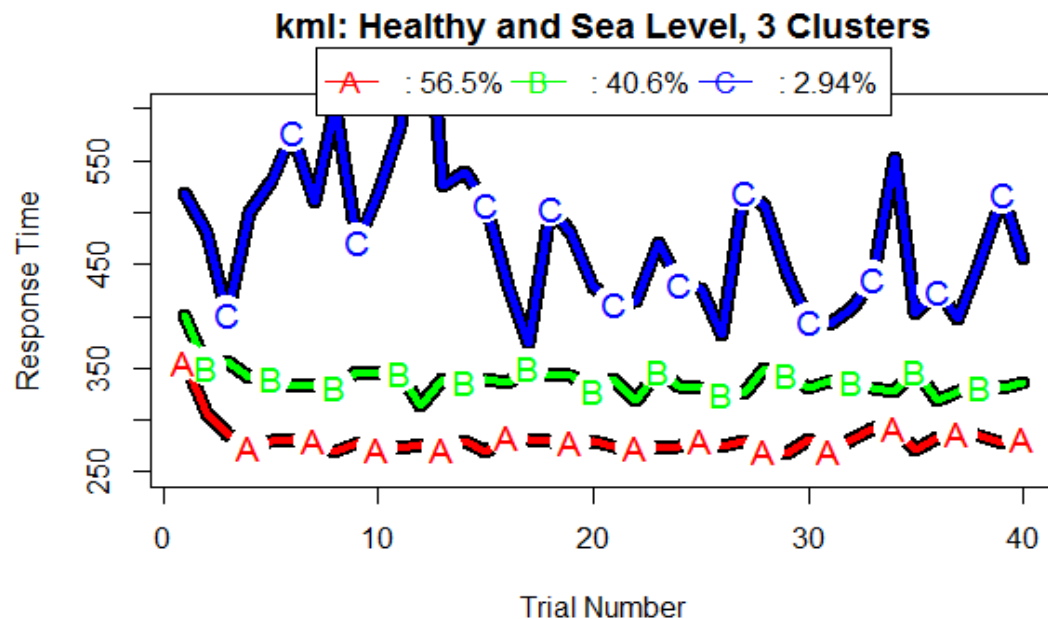
We return to the "known non-normals" to determine how frequently individuals from these groups are classified into cluster C based on the n=170 data set. Table 4 presents the number of each set of subjects that is classified into each cluster. The Euclidean distance results are very similar to those from the full set of data (n=389), with the exception of a slight decrease in the cluster A assignments of "non-healthy" Ft. Hood subjects. This indicates that the predictions are slightly better when excluding the Ft. Hood data from the cluster analysis because more "non-healthy" Ft. Hood subjects are classified into clusters B and C when the Ft. Hood data are excluded from the estimation of the k-means clusters (57.1% vs. 50.0%). Almost half of the "normal" Ft. Hood subjects are classified into cluster A and about 95% are in either cluster A or B. However, similar results hold for the "non-normal" Ft. Hood subjects.

Table 4: Out-of-Sample Class Predictions, 3 Clusters, Reduced Data						
	Manhattan Distance			Euclidean Distance		
Group	A	B	C	A	B	C
Above Sea Level	11 (52.4%)	8 (38.1%)	2 (9.5%)	8 (38.1%)	11 (52.4%)	2 (9.5%)
Concussed Cadets	3 (50.0%)	2 (33.3%)	1 (16.7%)	3 (50.0%)	2 (33.3%)	1 (16.7%)
Ft. Hood Healthy	123 (55.7%)	90 (40.7%)	8 (3.6%)	108 (48.9%)	102 (46.2%)	11 (5.0%)
Ft. Hood Non-Healthy	47 (48.0%)	49 (50.0%)	2 (2.0%)	42 (42.9%)	52 (53.1%)	4 (4.1%)
Healthy Seniors	0	9 (40.9%)	13 (59.1%)	0	9 (40.9%)	13 (59.1%)
Alzheimer Seniors*	0	6 (60.0%)	4 (40.0%)	0	6 (60.0%)	4 (40.0%)

* Based on only 25 trials.

GROUP-BASED TRAJECTORY MODELING

Group-based trajectory modeling is the second approach that can be applied in an unsupervised learning setting. This modeling approach allows for different models to be fit to each of the identified groups. It is based on a mixture of the models for each group, where "mixture" means that each individual belongs to a given group based on their trajectory and the overall probability of group membership. An advantage of this approach is that it allows for inclusion of covariates such as age, health history, etc. when determining group membership. As in the k-means clustering approach, it is possible to estimate the probability of group membership for any new subjects.

This approach was implemented using PROC TRAJ in SAS. Similar to results that were presented for k-means, the trajectory analyses forced clustering into two and three groups for each of the outcomes and subgroups examined.

SRT results using group-based trajectory modeling

Two sets of trajectory models were fit to the full data set of 389 "normal" subjects and then to the 170 normal subjects from the combined Air Force and altitude data sets. In each case, models contained a quadratic term for time and they assumed that SRT was approximately normally distributed.

Full data set analysis (N = 389 subjects)

Two-group analysis

The output below is from a model that assumed time follows a quadratic model and that the data cluster into two groups. There is a separate model presented for each of the two groups (yellow) with group 2 having a linear term that is larger in absolute value (-3.8988 v. -2.37423). Approximately 79% of subjects are assigned to group 1 and the remaining subjects are assigned to group 2.
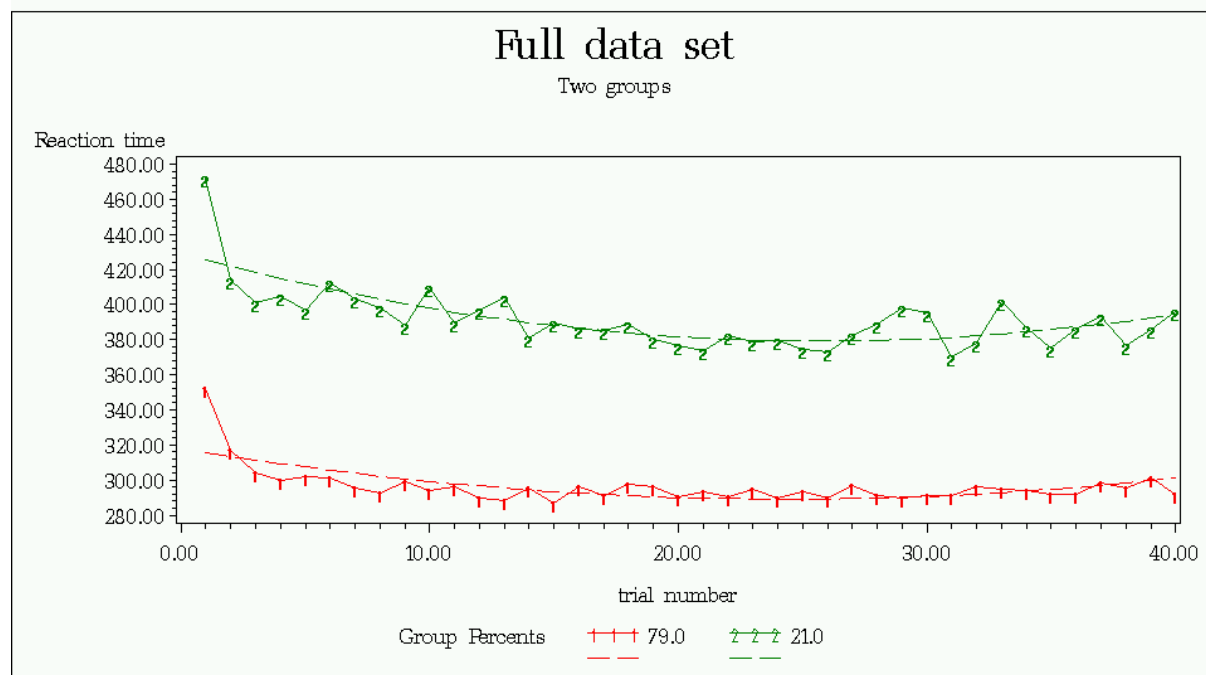
```
              Maximum Likelihood Estimates
              Model: Censored Normal (CNORM)
```

```
                                      Standard        T for H0:
Group    Parameter      Estimate         Error     Parameter=0     Prob> |T|

1        Intercept    318.06479       2.43337         130.710        0.0000
         Linear        -2.37423       0.25873          -9.176        0.0000
         Quadratic      0.04882       0.00611           7.993        0.0000

2        Intercept    429.38026       5.00765          85.745        0.0000
         Linear        -3.89880       0.50944          -7.653        0.0000
         Quadratic      0.07570       0.01203           6.293        0.0000

         Sigma         80.07444       0.45534         175.858        0.0000


         Group membership
1              (%)     78.96671       2.28546          34.552        0.0000
2              (%)     21.03329       2.28546           9.203        0.0000

BIC=-90319.69 (N=15528)   BIC=-90304.94 (N=389)   AIC=-90289.08   L=-90281.08
```

The plot below shows the estimated trajectories obtained from the analysis above. Note that these trajectories are very similar to those observed in the k-means cluster analysis that is described earlier in this report.

Plots of trajectories for the full data set (n=389) assuming two clusters

Three group analysis

The models that were fit for the three group analysis are identical to those for two groups, including is an assumption of a quadratic model with respect to time. Below is a summary of the model results from this analysis:

```
                  Maximum Likelihood Estimates
                   Model: Censored Normal (CNORM)

                                  Standard      T for H0:
Group    Parameter    Estimate      Error     Parameter=0    Prob> |T|

1        Intercept    308.65144    2.74523      112.432       0.0000
         Linear        -2.55731    0.29115       -8.783       0.0000
         Quadratic      0.05331    0.00691        7.717       0.0000

2        Intercept    376.64576    3.85345       97.743       0.0000
         Linear        -2.85110    0.38681       -7.371       0.0000
         Quadratic      0.05535    0.00925        5.984       0.0000

3        Intercept    499.31102   10.42941       47.875       0.0000
         Linear        -3.18731    1.13813       -2.800       0.0051
         Quadratic      0.06183    0.02780        2.224       0.0261

         Sigma         77.14375    0.43901      175.722       0.0000

         Group membership
1            (%)       59.83332    3.03212       19.733       0.0000
2            (%)       35.68415    2.93502       12.158       0.0000
3            (%)        4.48253    1.07623        4.165       0.0000

BIC=-89858.08 (N=15528)   BIC=-89835.96 (N=389)   AIC=-89812.18   L=-89800.18
```

In this case the AIC—a tool to assist with model selection—is smaller for the two group model which would favor the use of two groups; however, the values are not that different between the two models. The figure below presents the plots for the three group analysis. Once again, these results look very similar to those obtained using k-means cluster analysis that was presented earlier.

Plots of trajectories for the full data set assuming three clusters

Full data set
Three groups

Reduced Data Set: N = 170 subjects

Two-group analysis

The model presented is identical to the earlier one for two groups, with the exception that it generates groups using the reduced data set that excludes the Ft. Hood "normal." The output from this model is presented below for each group.

```
                     Maximum Likelihood Estimates
                     Model: Censored Normal (CNORM)


                                 Standard      T for H0:
Group    Parameter    Estimate      Error    Parameter=0    Prob> |T|

1        Intercept    315.93860    3.41204       92.595       0.0000
         Linear        -2.26969    0.37111       -6.116       0.0000
         Quadratic      0.04704    0.00878        5.357       0.0000

2        Intercept    425.53297   10.07005       42.257       0.0000
         Linear        -3.27872    0.90887       -3.607       0.0003
         Quadratic      0.05416    0.02131        2.541       0.0111

         Sigma         78.64789    0.67822      115.961       0.0000
```

```
        Group membership
1               (%)      84.69217        3.16776          26.736        0.0000
2               (%)      15.30783        3.16776           4.832        0.0000

BIC=-39249.70 (N=6768)  BIC=-39234.96 (N=170)  AIC=-39222.42  L=-39214.42
```

The plot corresponding to this analysis is presented below. Note that there is more variability in this plot when compared to the full data set.

**Plots of trajectories for the n=170 (Air Force and Altitude) data sets
assuming two clusters**
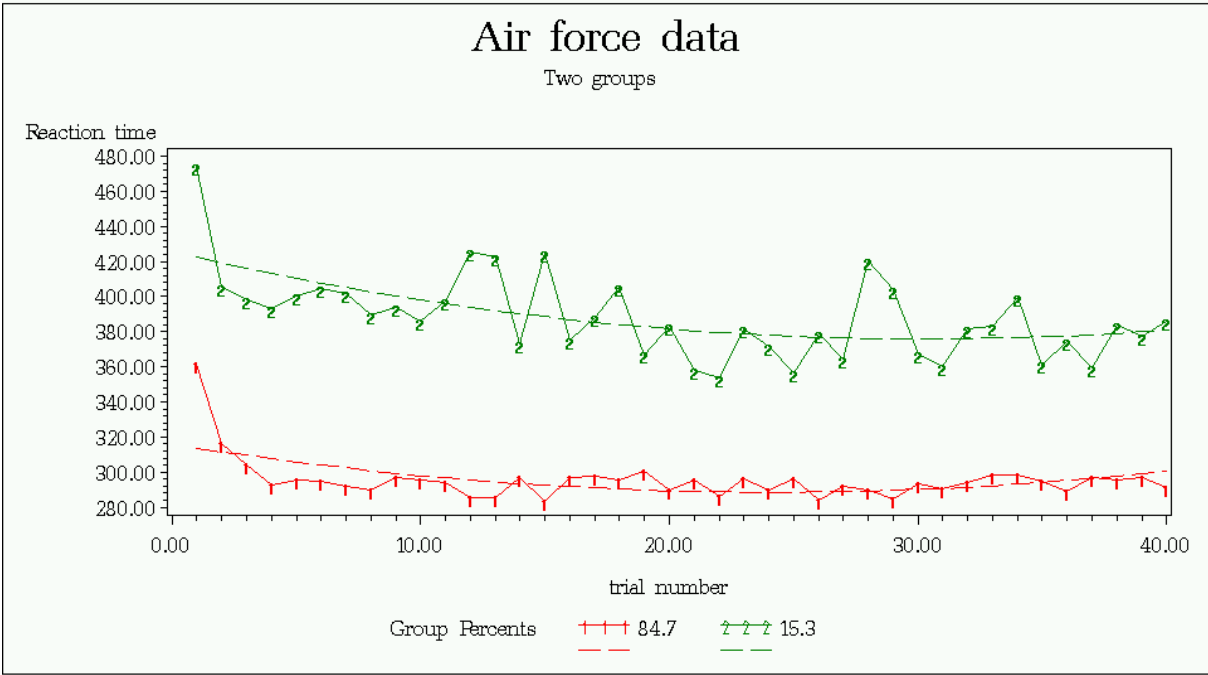
<u>Three group analysis</u>
The results presented below are for the three group analysis for the reduced (n=170) data set.

```
                    Maximum Likelihood Estimates
                    Model: Censored Normal (CNORM)

                                  Standard      T for H0:
Group   Parameter    Estimate       Error    Parameter=0    Prob> |T|

1       Intercept   303.26562     4.12403        73.536        0.0000
        Linear       -2.71330     0.44693        -6.071        0.0000
        Quadratic     0.05973     0.01049         5.693        0.0000
```

```
2        Intercept   356.21768    4.88209          72.964          0.0000
         Linear       -1.77742    0.53268          -3.337          0.0009
         Quadratic     0.03004    0.01248           2.407          0.0161

3        Intercept   532.91487   15.78672          33.757          0.0000
         Linear       -5.01212    1.76820          -2.835          0.0046
         Quadratic     0.06687    0.04227           1.582          0.1137

         Sigma        75.36235    0.64995         115.951          0.0000

         Group membership
1            (%)      56.08676    4.17946          13.420          0.0000
2            (%)      40.40807    4.13598           9.770          0.0000
3            (%)       3.50518    1.41833           2.471          0.0135
```

BIC=-39034.68 (N=6768)   BIC=-39012.57 (N=170)   AIC=-38993.76   L=-38981.76

In this case, the AIC values are very close, indicating little difference between the two- and three group models. The plots of the trajectories are presented in the figure below.

Plots of trajectories for the n=170 data set, assuming three clusters
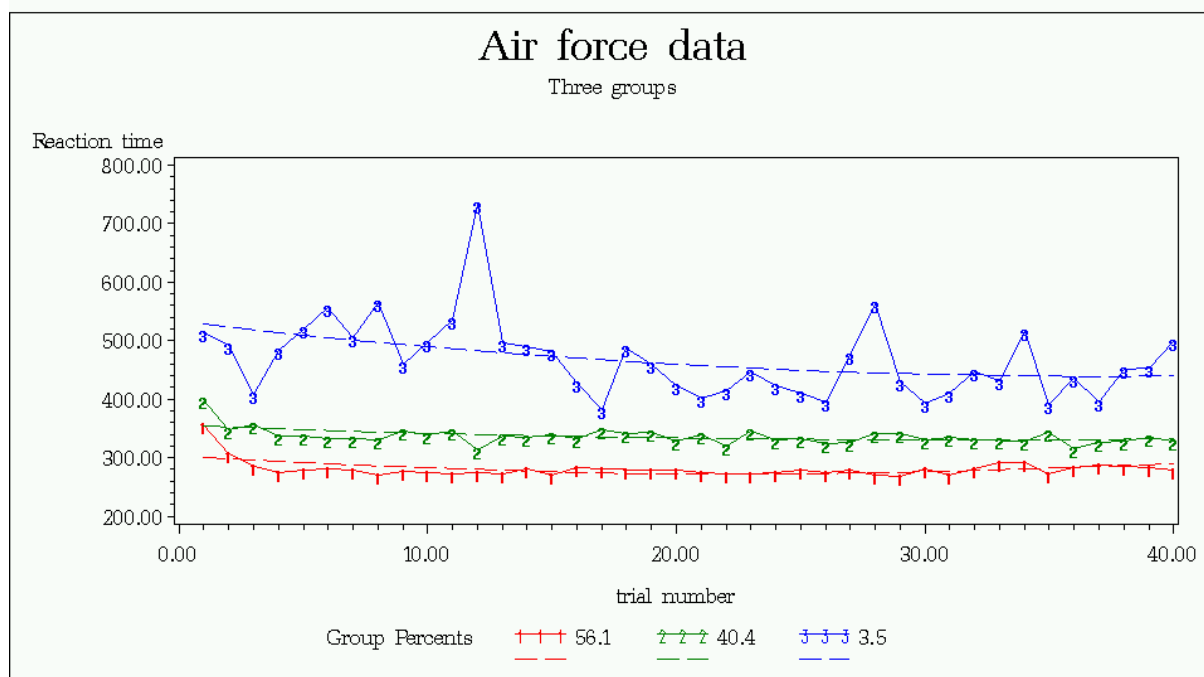
SUMMARY

kml vs. *PROC TRAJ* Classification Comparison

k-means clustering and group-based trajectory modeling rely on different assumptions and different statistical tools. However, these results show that clustering of "normal" subjects using the two methods is very similar. In fact, of the 389 subjects in the full "normal" dataset, we find 97% and 90% agreement in cluster assignment for the 2 cluster and 3 cluster analyses, respectively. Table 5 presents the cross-tabulation of cluster assignment for the two methods.

Table 5: PROC TRAJ vs. kml "Healthy" Subjects Classification, Full Data					
	PROC TRAJ 2 Cluster		PROC TRAJ 3 Cluster		
kml Cluster	1	2	1	2	3
A	294	0	205	0	0
B	12	83	29	130	0
C	-	-	0	8	17

When trajectory modeling and k-means disagree, k-means assigns the subject to the next higher mean response time group. Importantly, because the cluster classifications are so similar between trajectory modeling and k-means, the trial-by-trial means for each cluster are similar. This results in only minor differences in out-of-sample predictions.

Repeated Measures Analysis: Phase V Report
Summary of results for the prediction of healthy in the Ft. Hood data set

We ran several different types of analyses with the goal of developing a classification rule for "normal/abnormal" within the Ft. Hood data set. The following statistical approaches were used:

- Logistic regression with health status as the outcome
- Group based trajectory modeling that included covariates
- Mixed model regression analysis with reaction times over the course of the 40 trials as the outcome

The results of these analyses highlight the need to either define the outcome of "healthy/unhealthy" more carefully or to proceed with the unsupervised learning methods and develop an approach that can be used to better ascertain the usefulness of the groupings that are identified as part of this analysis. One additional important finding arose from the mixed model results and pointed to the fact that much of the information over the course of the tests was available in the first 20 -25 trials and will be discussed further below.

Logistic regression analyses

We fit a series of logistic regression models to the outcome "healthy/unhealthy" and focused on the development of a series of summary measures for the 40 trials of simple reaction time. In total, seven different summary measures were considered; the mean reaction time over the 40 trials (MEAN), the median reaction time of the 40 trials (MEDIAN), the standard deviation of the reaction time over the 40 trials (SD), the minimum reaction time over the 40 trials (MIN), the maximum reaction time over the 40 trials (MAX), the difference between the maximum and minimum reaction time (DIFF), and the percent difference between the maximum and minimum reaction time computed as 100 * (maximum reaction time – minimum reaction time)/minimum reaction time (%DIFF). A series of logistic regression models were then fit with each of these summary measures and additional covariate for age and gender. The area under the receiver

operating characteristic curve (ROC) was then computed for each measure and is reported in the table below. A ROC value of 0.5 is equivalent to using a fair coin toss for the determination of group membership. Note that the largest ROC value was 0.59. Fitting a model with more variables did not increase this value. In general the "best" summary measures were the maximum reaction time over the course of the 40 trials and the difference between the maximum and minimum response time. Note also that age and gender were better predictors of group membership than the summary measures from the test.

Table 1. **Area under the receiver operating characteristic curve for predicting the "healthy/unhealthy" outcome based on a logistic regression model**

Variables included in the model	Area under the receiver operating characteristic curve
Age and gender	0.57
Simple reaction time summary measures	
MEAN	0.53
MEDIAN	0.53
SD	0.53
MIN	0.49
MAX	0.54
DIFF	0.54
%DIFF	0.54
Age and gender coupled with SRT measures	
AGE, GENDER, MEAN	0.58
AGE, GENDER, MEDIAN	0.58
AGE, GENDER, SD	0.58
AGE, GENDER, MIN	0.57
AGE, GENDER, MAX	0.59
AGE, GENDER, DIFF	0.59
AGE, GENDER, %DIFF	0.58

Group-based trajectory modeling

We ran a series of group-based trajectory models that included the following covariates: health status, age group, and gender. While addition of these covariates changed group membership slightly, they had little effect on the results that were previously presented.

Mixed models of simple reaction time

We fit a series of mixed models to the simple reaction time data with the outcome being the value of simple reaction time and the covariates including the following: time, group membership (healthy/unhealthy) and an interaction term for time by group membership. These models were different from those fit in earlier analyses as time was treated as a class variable so the model did not assume any functional form for time. The results obtained from this modeling exercise were very interesting in that time was not statistically significant after the $20 - 25$ th trial.

Potential next steps

The results of these analyses pointed out several key results:

1. Summary measures of the 40 trials of simple reaction time are not predictive of "healthy/unhealthy".
2. Mixed models demonstrated that much of the information is in the early segments of the 40 trials with information trailing off after the 25th trial.
3. Group-based trajectory modeling and cluster analyses can be used to identify three groups with the third group containing fewer than 10% of the subjects and generally having the longest reaction times.

Based on these results next steps can include:
1. Refine the definition of "unhealthy" to make it more restrictive.
2. Focus on unsupervised methods applied only to the sub groups of healthy and unhealthy to see if these methods identify groups that may be of further interest.
3. Include results from other tests in the analysis. This can be done first with the logistic regression approach as well as the unsupervised approaches.

Additional analyses should be well-planned with care given to the best overall approach of supervised vs unsupervised learning coupled with a methodology to better identify "unhealthy". For example, an unsupervised approach can be used to create "rules" for identifying "unhealthy" individuals in a larger data set. These rules can then be applied to a smaller set of subjects who have a more extensive health assessment and a better refined definition of "unhealthy" to assess the overall usefulness of the battery of tests in this setting.

Summary of trial-by-trial-level analyses

Although much has been published on summary statistics of multi-trial cognitive function tests, little has been done to leverage all the trial data upon which these summary measures are based. It may be possible to use these data in ways that can improve on current strategies to identify people with head injury, depression, PTSD, and age-related cognitive decline. Is possible, these strategies could have direct application to mission readiness among military personnel. This document summarizes work that has been done by AnthroTronix to explore how the repeated measures that are collected during computerized cognitive testing might be used to design new ways to identify various types of clinically relevant abnormalities that inform on mission readiness. This work used multiple data sets – both military and civilian – to pursue this line of investigation.

Among the most challenging aspects of this work was to use repeated measures to identify people who are within and outside the "normal" range of values. Some of our early results on simple reaction time are presented below for young, healthy individuals who received cognitive testing at sea level and at extreme altitude. By using simple trial-specific means in combination with smoothing and modeling techniques, we showed that, among the same individuals who were testing in different settings (sea level and altitude), repeated measures of reaction time at altitude were less favorable and did not stabilize over time as they did among the same individuals at sea level.
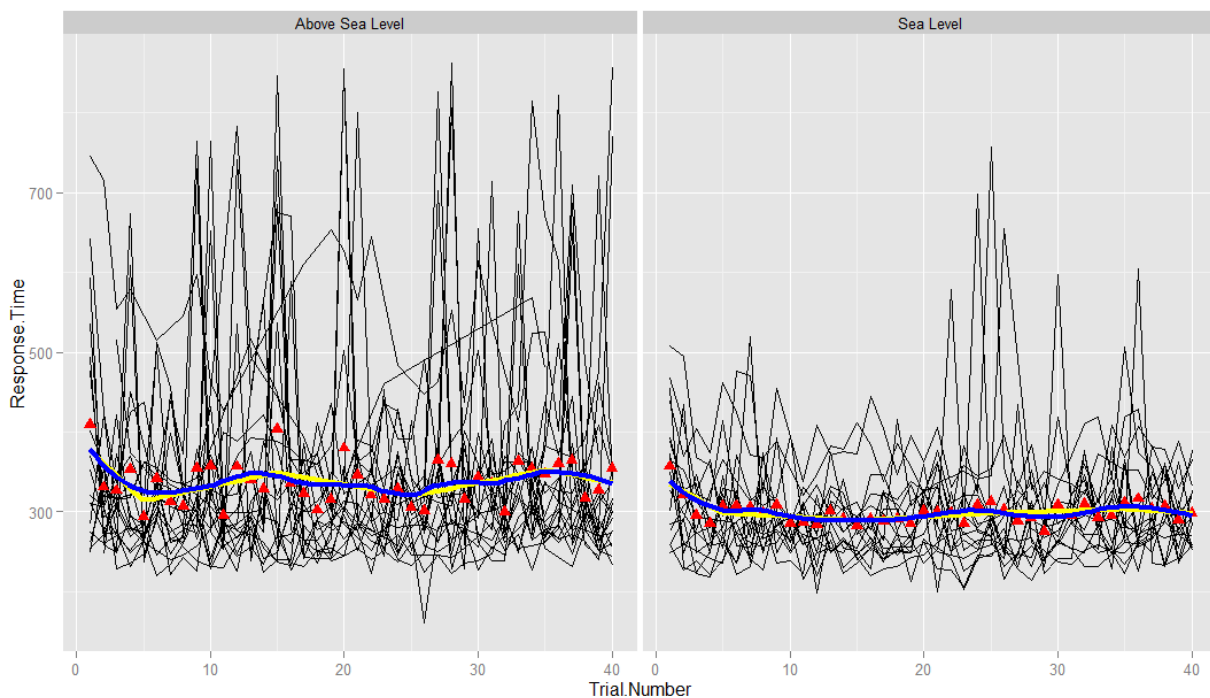
These results showed differences in the shape of the curves over time between the two testing conditions and led us to extend this work to other data sets to determine if this line of inquiry continued to hold promise. In a related set of analyses, we used data collected from Air Force Academy cadets, some of whom reported concussion, to examine how our earlier analyses could be applied in the setting of head injury.
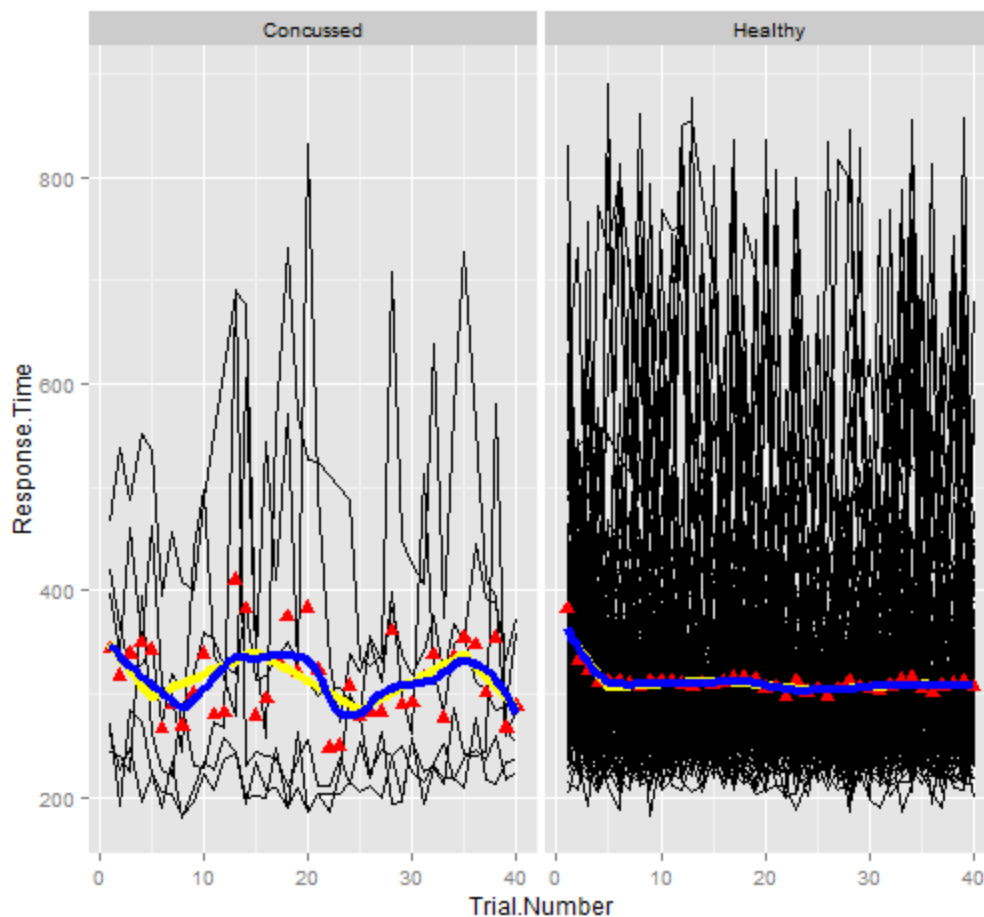
Despite the relatively small number of cadets reporting head injury, our results were remarkably consistent with findings from the altitude data set. The differences in the shape of the average response time trial profile suggest the possibility of a tool for distinguishing non-normal subjects from normal subjects.

Because the typical concussed subject's trial profile was somewhat erratic, it may be possible to implement a smoother at the subject level to pick up any distinct shape in the profile as compared to the average trial profiles. These techniques are common in functional data analysis. Mixed models also offer a possibility for distinguishing non-normal subjects from normal subjects. The random intercept model implemented for these analyses produces estimates of subject-specific shifts from the average response time. It may be useful to extend these models to incorporate random slopes, which produce subject-specific estimates of changes in slope from the average changes in slope. Further investigation into the utility of subject-specific random intercepts/slopes and subject-level smoothers is well suited to efforts aimed at identifying non-normal subjects.

Beyond these concepts, we were also interested in whether these approaches could be used with cognitive tests other than reaction time.
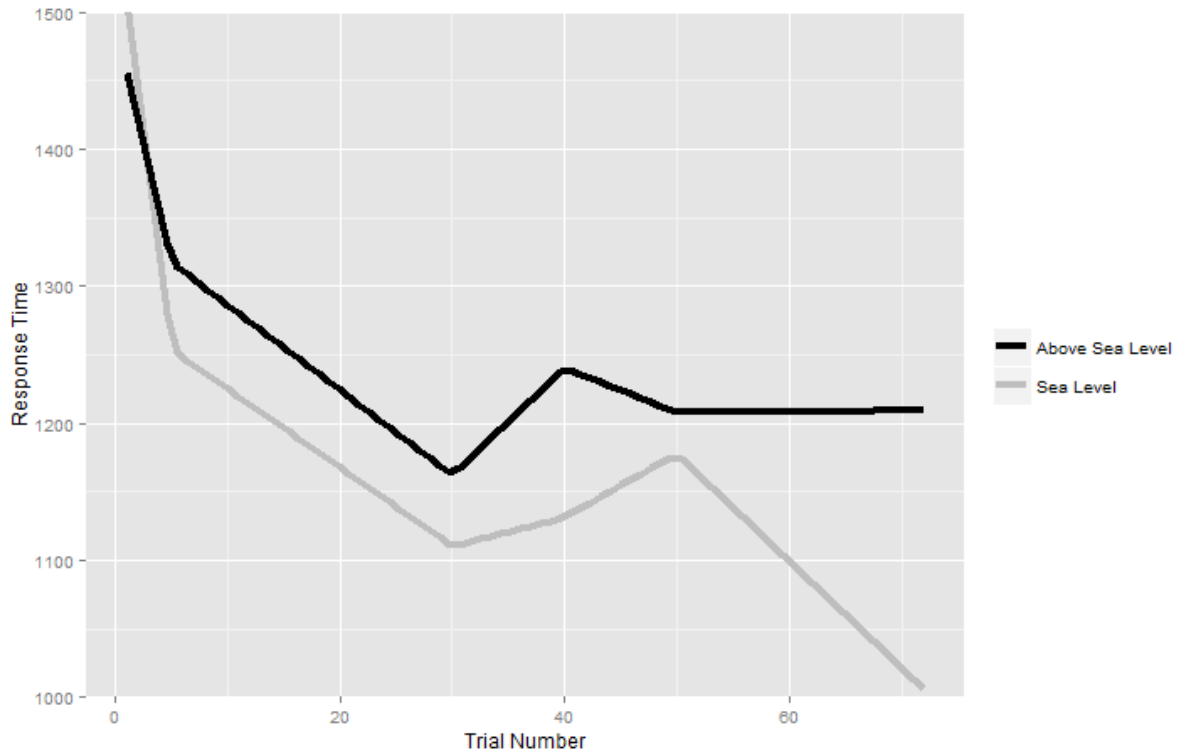
The figure above summarizes code sub-learning results from young, healthy individuals in the altitude data sets. The figure shows that for this test, hypoxia results in less favorable results over time, with a notably less favorable learning response in later trials.

Additional work extended these analyses to data collected among active duty personnel at Ft. Hood, and to older adults. The objective was to define a classification scheme in which response patterns could be used to develop a reasonably precise tool to distinguish normal from non-normal. In practice, it is often a challenge that subjects are not predefined as "normal" or "abnormal." This leads to the need to look for groups within a data set that are similar in behavior over the course of a given longitudinal trajectory. We selected strategies that rely on statistical methods that fall in the category of unsupervised learning techniques. This is a type of machine learning algorithm that is used to draw inferences from datasets consisting of input data without labeled responses (e.g. normal or non-normal). These approaches were used for exploratory data analysis to find hidden patterns or groupings in data. With this approach, the current state (e.g., normal, abnormal) of the individual is not used to identify predictors of that state; rather the data are divided in a systematic manner into groups that behave similarly.

kml: Healthy, Sea Level & Ft. Hood, 3 Clusters

A : 53.7% B : 41.9% C : 4.37%
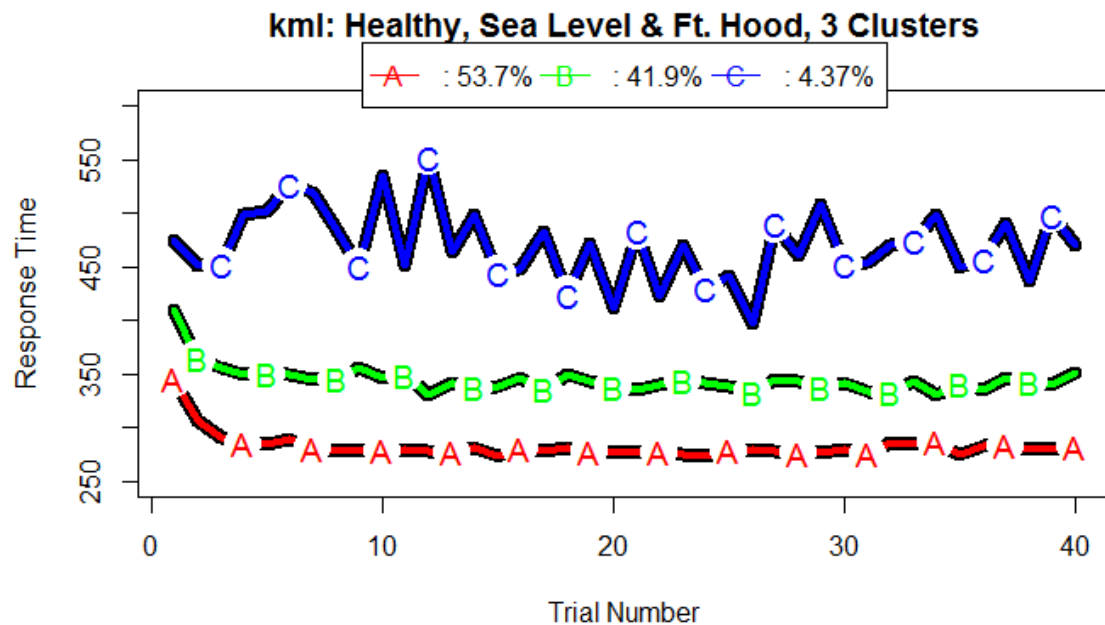
An example of this work is shown above, in which data from Ft. Hood and the altitude data set were combined and used to identify clusters of response trajectories that fell into natural groups that could then be used to distinguish normal from non-normal. In this example, about 95% of subjects fell into two clusters that had relatively smooth response patterns, one of which had a higher mean than the other, and the remaining 5% of subjects fell into a group with an even higher mean and unstable response pattern. We examined this statistical approach using two distinct strategies, and these yielded remarkably consistent results, as shown in the figure below.

Full data set

Three groups

Reaction time

Group Percents | +++ 59.8 | 2-2-2 35.7 | 3-3-3 4.5
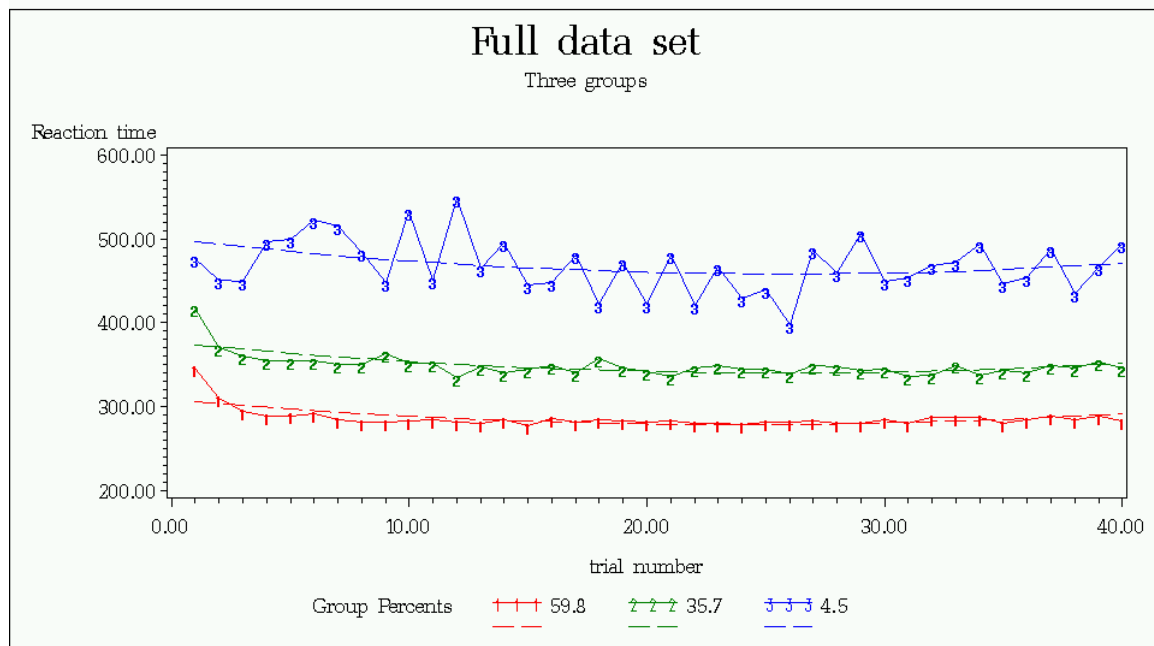
We also applied the *k*-means clustering technique to a sample of college athletes. DANA was administered to all athletes for preseason baseline testing. Then, all participants were followed over the course of the season, and if they sustained a concussion, they were re-administered DANA within 24 hours of their head injury. The figure below shows trajectories for baseline testing on the Simple Reaction Time 1 subtest.

Concussion Data: SRT1

Legend: A : 66.7%   B : 31.3%   C : 1.99%

Y-axis: Response Time (200, 400, 600, 800)
X-axis: Trial Number (0, 10, 20, 30, 40)
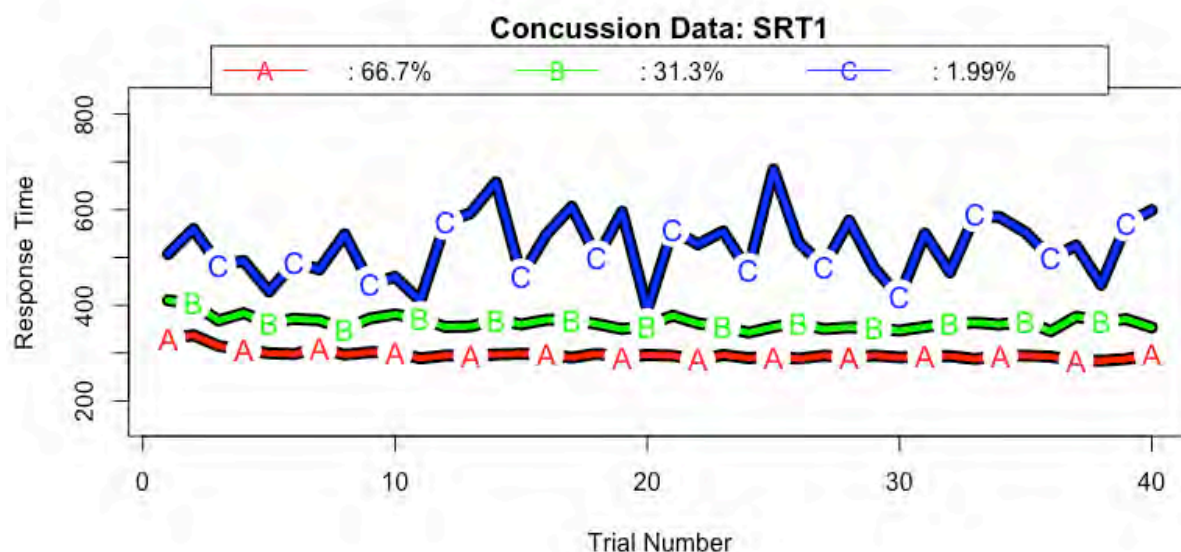
Similar to the analyses reported above, the three clusters identified correspond to a relatively fast and stable group (A), a somewhat slower and more variable group (B), and finally, a much slower and much more variable group (C). The clustering results for the Simple Reaction Time 2 subtests were similar as shown below:

Concussion Data: SRT2

Legend: A : 62.7%   B : 31.3%   C : 5.97%

Y-axis: Response Time (200, 400, 600, 800)
X-axis: Trial Number (0, 10, 20, 30, 40)
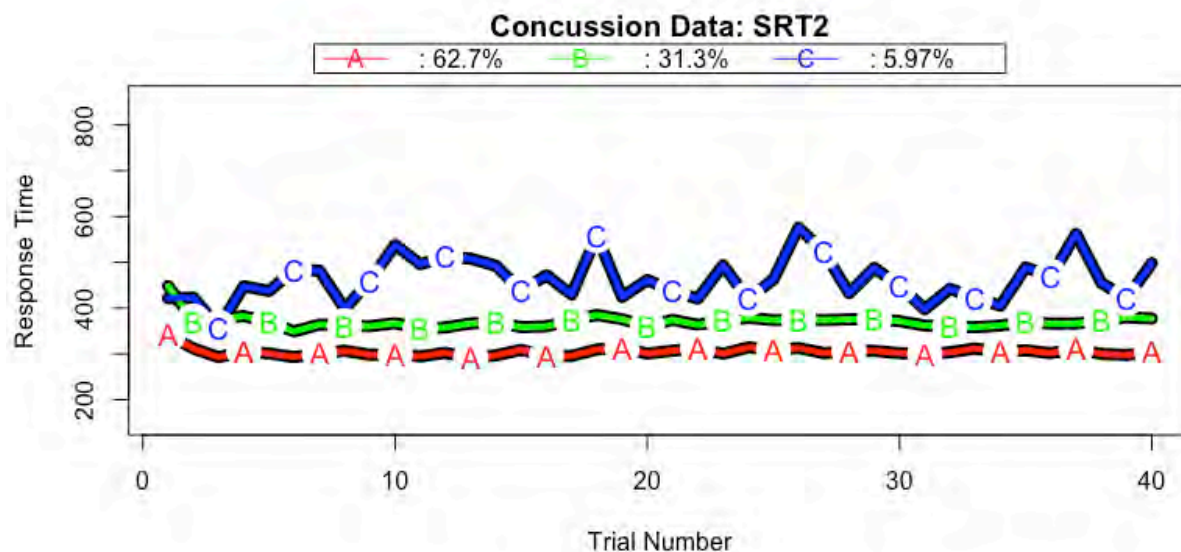
These clusters can be used to categorize new trajectories from new subjects. For example, if a subject is suspected of having a concussion and their post-injury response time trajectory is most similar to group C, the slowest and most variable group, then this subject's performance would be unusually poor, suggesting further follow-up.

2

3

4

Trial-by-Trial Analysis: A New Approach to Detecting Neurocognitive Changes With Computerized Cognitive Tests

Darcy Steeg Morris, PhD, Corinna Lathan PhD, Timothy Lacy, MD, and Helaine E. Resnick, PhD, MPH*

Darcy Steeg Morris, PhD, Resnick, Chodorow and Associates, Silver Spring, MD
Corinna Lathan, PhD Anthrotronix, Inc, Silver Spring, MD
Timothy Lacy, MD Anthrotronix, Inc, Silver Spring, MD
Helaine E. Resnick, PhD, MPH, Resnick, Chodorow and Associates, Silver Spring, MD

*Corresponding Author:
Helaine E. Resnick, PhD, MPH
8737 Colesville Road
Silver Spring, MD 20910
heresnick@gmail.com
tel: 202.329.8616

23

24

29

30

31

32

33

34

35

36

37   ABSTRACT

38   Computerized cognitive testing quantifies performance by generating a mean from multiple trials

39   of a given test, such as simple reaction time (SRT). This report takes advantage of the richness

40   of trial-by-trial SRT data to explore a method that distinguishes groups from each other based on

41   the pattern, rather than the mean of their responses. Using two data sets that include subjects

42   with altitude-induced hypoxia and concussion, as well as study-specific controls for these

43   exposed subjects, we fit loess curves that provided a graphical summary of the relationship

44   between response time and trial number. Based on these shapes, we used spline regression to fit

45   models to localized areas of group-specific curves, and then used these models to determine

46   whether the slopes of the curves differed between groups at various points across 40 SRT trials.

47   We observed significant differences in the slopes of SRT response curves between concussed

48   and non-concussed individuals and among individuals at sea level and then at extreme altitude.

49   Differences in the patterns of these curves suggested less favorable SRT performance among

50   concussed and hypoxic subjects despite the similar mean SRT among concussed and non-

51   concussed individuals. We present the clinical implications of further developing this method to

52   evaluate an individual's response pattern against a normative pattern for the purposes of

53   detecting cognitive deficit.

54

55

Introduction

Cognitive testing has many applications, including screening, patient care, and drug development.[1-6] Cullen's review[1] of thirty-nine cognitive screening tools highlights the breadth of available testing strategies, as well as some of the challenges associated with using these tools for differential diagnosis and longitudinal assessment. Cognitive testing batteries often evaluate numerous functions (e.g. memory, visual construction, reasoning, etc.), and in many cases, subtests involve multiple trials that are collapsed into a summary score, often a mean. In turn, these summary scores—used alone, or in combination with summary scores from other cognitive subtests—are used to globally assess cognitive health and progression of cognitive decline over time. This aggregated approach assumes that (1) a mean provides an accurate reflection of response times across trials, and (2) no clinically relevant information can be gleaned from the *shape* of an individual's response curve across the trials.

With these assumptions in mind, summary scores are often collapsed into categorization schemes that place cognitive status in a binary (impaired/not impaired) or ordinal (normal/mild impairment/moderate impairment/severe impairment) evaluation framework. This is true for traditional neuropsychological cognitive batteries, brief neurocognitive screening tools, as well as for both "pen and paper" scoring and some computerized applications.[1,6-8] The traditional practice of using binary or ordinal scoring frameworks to detect cognitive deficits has important limitations. First, this approach is constrained by the maxim value dictated by the sum of scores on the component tests. Ceiling effects have been shown to be problematic in a number of settings, and these limitations can become exacerbated in longitudinal research.[9] A second limitation concerns a scenario in which pooling sub-scores can mask cognitive impairment on one subtest while still yielding a favorable overall score. A third limitation involves the assumption that underpins use of mean scores as a tool to summarize a series of individual trials: Using means to summarize a series of measures assumes that the clinical utility of the measures lies entirely in a single overall score, rather than in potentially informative patterns of fluctuation over the course of many trials.

A longstanding reliance on summary measures has resulted in failure to optimize use of information that is readily available from individual trials in many computerized cognition tests.

87 For example, simple reaction time (SRT) assesses psychomotor speed—often in response to a

88 visual stimulus. This test can generate numerous summary variables including: the number of

89 early responses; the number of trials in which no response occurred; number of completed trials;

90 mean of completed reaction times, and standard deviation of completed reaction times.[10] While

91 older reaction time studies involved as many as 100 trials,[11] newer ones often have between 20

92 and 50.[1,10,12,13]   Regardless of the number of trials, SRT summary scores have been used as a

93 means to describe an individual's global performance on this subtest without regard to

94 quantifying the shape of the curve that is generated by performance on each trial.

95

96 Against this backdrop, new technologies allow capture, export, and analysis of computerized

97 cognitive testing data in a manner that facilitates examination of trial-by-trial data, potentially

98 unlocking applications of this information that are new, highly quantitative, and clinically

99 relevant.   Methods to quantify subtle patterns across a series of cognitive subtest trials could

100 reveal previously-unidentified deficits and perhaps the etiology of some forms of cognitive

101 dysfunction.   We describe one such method and apply it to two data sets: a study of college

102 students who engaged in cognitive testing at sea level and at extreme altitude, and a study of

103 concussed and non-concussed college athletes.

104

105 **Materials and Methods**

106 DANA is a hand-held, FDA-cleared clinical neurocognitive assessment tool that measures and

107 tracks changes in cognitive efficiency by measuring response speed and accuracy.[14] DANA

108 includes eight cognitive tests and seven psychological questionnaires that measure multiple

109 aspects of brain health. DANA has been validated in diverse military and civilian research

110 settings.[14-16] It assesses reaction time by measuring the time between when an on-screen

111 stimulus is triggered and when it records either capacitance or force on the screen. DANA's

112 subtests (SRT, code substitution, procedural reaction time, spatial processing, go/no go, and

113 memory search) include multiple trials within each testing protocol, and DANA records and

114 timestamps the response input for each trial. The SRT subtest consists of 40 trials. Subjects are

115 presented with a stimulus in the center of the screen and asked to respond as quickly as possible.

116 DANA tracks whether a participant did not respond on a given trial ("lapsed" trials) or

117 responded too quickly ("fast" trials). Both of these responses are potential sources of

118 measurement error, and they are reported separately for each subtest. Using the trial-by-trial

119 response time data, DANA calculates summary measures for each participant. These include

120 mean, median, and standard deviations of all reaction times for each subtest. We focus on SRT

121 because our previous work showed that it is sensitive to hypoxic impairment, and this test has

122 also been used to identify deficits in Alzheimer's disease, mild cognitive impairment,

123 Parkinson's disease, depression, and insomnia.[16-22]

124 This report uses DANA's trial-by-trial SRT data from two studies. The first administered DANA

125 to the same individuals at sea level and at extreme altitude (the "altitude data set"), and the

126 second study administered DANA to two groups of college students, one with, and the other

127 without concussion (the "concussion data set"). The altitude study has been described.[23]

128 Briefly, 21 healthy, physically active subjects (12 males and 9 females, average age 20.8 yrs,

129 range 19–23 yrs) were studied first at 130m and then again at 5,260m. The concussion study

130 involved 159 college students, 6 of whom had self-reported concussion.[15] The altitude study was

131 performed according to the Declaration of Helsinki and was approved by the Institutional

132 Review Boards of the University of Colorado and the University of Oregon, as well as the

133 Human Research Protection Office of the US Department of Defense. Data for the concussion

134 study were collected under the U.S. Air Force Academy performance improvement protocol.

135 The objective of this exploratory study was to take advantage of the richness of the SRT data to

136 develop a method to study group differences that incorporates the trial-by-trial dimension of the

137 test. We began by categorizing subjects in each data set according to whether they were in a

138 "pre-exposure" state (i.e. at sea level or no reported concussion) or if they had an exposure that

139 was hypothesized to unfavorably impact cognitive function (i.e. altitude-induced hypoxia or self-

140 reported concussion). We then plotted SRT response times for each subject in both data sets,

141 stratified by exposure status. Using these plots as guides, we sought to identify a statistical model

142 that captured trends in the trial-level mean response times and a strategy to capture differences in

143 both the means and variability of trial profiles.

144 To achieve this goal, we fit loess curves to each data set to visualize the shape of an "ideal"

145 smooth curve. Loess curves are a type of non-parametric smooth fit that is data-driven and

146 empirically derived. They do not require an *a priori* model specification, nor do they rely on

147   parametric assumptions about the shape of the trend.[24] This technique estimates a curve by

148   fitting multiple simple models to localized ranges of the x-axis, and it provides an appealing

149   graphical summary of the relationship between response time and trial number. This approach to

150   modeling the mean response function uses information of the surrounding trials to define a

151   smooth and visually intuitive trend. This is accomplished without losing important features of

152   the data. The loess requires setting a bandwidth parameter that controls the smoothing. We

153   choose a bandwidth of 0.35 (35% of data points are included in the span used for the local

154   regressions) that visually balanced over- and under-fitting. The loess curves for both the altitude

155   and concussion data sets indicated local curvature across trials.

156   Using the loess curves as a target for regression modeling, we first fit quadratic models to

157   capture the changing shapes of the loess curves, but this strategy did not provide adequate fit.

158   This suggested that linear spline regression might be better candidate for capturing localized

159   curvature across trials. Spline regression[25] provides piecewise fit in which a set of separate

160   linear models are fit in localized areas and joined together to estimate a curve. For ease of

161   interpretation, we used truncated linear splines of degree one.

162   The basis of truncated linear splines is provided by using explanatory variables with the
163   following form:

$$(\text{Trial Number} - \kappa_k)_+ = \begin{cases} \text{Trial Number} - \kappa_k & \text{if Trial Number} > \kappa_k \\ 0 & \text{otherwise} \end{cases}$$

164   where $\kappa_1, \dots, \kappa_K$ is a set of "knots" - points at which two separately sloped lines join together.

165   The full linear spline regression equation for each condition is:

$$E(\text{Response Time}|\text{Trial Number})$$

$$= \beta_0 + \beta_1 * \text{Trial Number} + \sum_{k=1}^{K} \beta_{1+k} * (\text{Trial Number} - \kappa_k)_+ \qquad (1)$$

166   An advantage of the first degree truncated linear spline is its ease of interpretation. In these

167   models, the estimated coefficients of the spline variables are interpreted as the additional slope

168   effect for a given range of trials. Using an example of 15 trials with a knot at trial number 5, β_1

169   (the coefficient associated with trial number) is the slope from trial number 1 to trial number 5,

170   and the additional slope effect from trial number 6 to trial number 15 is β_2. The total slope from

171   trial number 6 to trial number 15 is $\beta_1 + \beta_2$. A higher order spline may better capture some more

172   of the localized curvature, but at the expense of interpretability.

173   The linear spline regression in equation 1 assumes that all observations are independent;

174   however, independence is not the case in the repeated measures data that are collected in

175   computerized cognitive testing. Thus, correlation among response times within a subject needs

176   to be taken into account. Linear mixed models are one tool that accounts for the correlation with

177   a trial-constant, subject-specific response time effect. Accordingly, we introduced a random

178   intercept into the linear spline model that allowed each subject's intercept to be different from the

179   others. Equation 1 is extended to include a random intercept as follows:

$$E(\text{Response Time}|\text{Trial Number})$$
$$= \beta_0 + \beta_1 * \text{Trial Number} + \sum_{k=1}^{K} \beta_{1+k} * (\text{Trial Number} - \kappa_k)_+ + u_i \qquad (2)$$

180   where *i* denotes the subject index and u_i is the random intercept.

181

182   **Results**

183   *Data Set Descriptions*

184   Table 1 provides summary information on trial-by-trial SRT data for the two data sets. Although

185   each subject completed 40 trials, some of these trials were discarded because they were either

186   lapsed or fast. The average number of valid trials was 37.4 and 39.8 among subjects at high

187   altitude and sea level, respectively. The corresponding average trial numbers for concussed and

188   non-concussed subjects were 39.3 and 39.8, respectively. In the altitude and concussion data sets,

189   there were 1,462 and 6,327 observations uniquely identified by subject, trial number and

190   exposure.

191

192   Compared to sea level, at high altitude, there was a markedly higher mean SRT response time

193   (337.1 vs. 299.0). This difference of -38.1 (p-value = 0.00) was statistically significant based on

194   a standard pooled t-test. Although concussed subjects also had higher mean response times

195   (314.8) than their non-concussed counterparts (310.7), this difference of -4.1 (p-value = 0.50)

196   was not statistically significant. Thus, using means as a summary statistic to describe SRT in the

197 two studies yields divergent findings concerning the contributions of exposure to high altitude

198 and concussion on SRT performance.

Plots of Individual and Mean Trial Responses

200 Figures 1 and 2 plot SRT response times for each subject in the altitude and concussion data sets,

201 according to exposure status. Each line represents one subject's response times and the red

202 circles depict trial-level mean response times. The patterns of trial-level mean response times

203 (with respect to both variability and absolute levels) illustrate the potential to uncover interesting

204 and significant differences in the shape of mean response time patterns for individuals under

205 "normal" and abnormal (i.e. hypoxia and concussed) conditions. Figure 1 shows that subjects'

206 mean response times at sea level were fairly even after what appears to be a brief learning period

207 in the first 5 trials. Under hypoxic conditions, subjects' mean SRT response times do not smooth

208 out over the course of 40 trials, remaining erratic throughout the test. A similar pattern is

209 observed in Figure 2, although some of the variability among concussed subjects may be

210 associated with their small numbers.

211

Visual Inspection of Loess Curves

213 Figures 3 and 4 show trial-level mean response times with corresponding loess curves and spline

214 regression fits for the altitude and concussion data sets. The loess curves indicate localized

215 curvature that is particularly evident among the hypoxic and concussed subjects relative to the

216 non-exposed state. The loess curve of the mean response time for subjects at sea level is around

217 300ms across all trials after trial number 5, while the corresponding curve for the hypoxic

218 subjects not only has a higher mean across trials than sea level, but it exhibits more variable

219 behavior (Figure 3). In addition to differences in means across the trials, mean response times

220 for each trial are more variable around the loess curve for the hypoxic subjects. After about trial

221 number 5, the sea level subjects' average profiles are strikingly linear, while curvature remains

222 evident among the above sea level subjects' average profiles.

223

224 Similar to sea level subjects in Figure 3, the loess curve of the mean response time for non-

225 concussed subjects is stable and consistently lies just above 300 with the exception of the first

226 few trials (Figure 4). In contrast, the corresponding curve for the concussed subjects fluctuates

227 above and below the curve for non-concussed subjects. Although there is more curvature in the

228 trial averages for concussed subjects, both groups have a similar trial-level average response

229 times: 311 and 314, respectively. The trial-level response times are more variable around the

230 loess curve for concussed subjects than for their non-concussed counterparts, an observation that

231 is likely due to the small number of concussed subjects. Regardless of the variability about the

232 trend, the curve exhibits pronounced humps that arise from changes in slope roughly around

233 trials 5, 15, 20, 25 and 35. Similar to sea level subjects in the altitude data set, non-concussed

234 subjects' average trial profile is strikingly linear after about trial 5.

235 ***Fitting Models to Curves***

236 Using the loess curves for hypoxic and concussed subjects as visual guides, the shapes of the two

237 curves suggest that changes in slope occur about every 5 trials. Accordingly, we defined knots at

238 trial numbers 5, 15, 25, and 35. Figures 3 and 4 show the linear spline model fit for the two data

239 sets. This model aligns very closely with the loess curves despite the models' simplifying

240 assumption.

241

242 Table 2 presents estimated coefficients, standard errors, and statistical significance from the

243 linear spline mixed models fit to the altitude and concussion data sets in Figures 3 and 4, by

244 hypoxia and concussion status. In the altitude data, there is a statistically significant negative

245 slope between trials 1 and 5 ($\hat{\beta}_1$ = -9.20) among sea level subjects. The magnitude of the slope is

246 statistically significantly different and slightly negative from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2$ = -9.20+8.23

247 = -.97). The slope does not significantly change after trial 15. For the hypoxia model, there is

248 also a statistically significant negative slope between trials 1 and 5 ($\hat{\beta}_1$ = -14.86). The magnitude

249 of the slope is statistically significantly different and positive from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2$ =

250 4.01), and continues to significantly change until trial 35. It is negative from trials 16 to 25

251 $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ = -3.47), positive from trials 26 to 35 ($\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4$ = 3.00), and does not

252 significantly change after trial 35. The features estimated from the linear spline mixed model are

253 consistent with patterns that are observed in Figure 3.

254 In the concussion data set, there is a statistically significant negative slope between trials 1 and 5

255 ($\hat{\beta}_1$= -13.90) for non-concussed subjects. The magnitude of the slope is statistically significantly

256 different and slightly positive from trials 6 to 15 ($\hat{\beta}_1 + \hat{\beta}_2$ = .78). The slope changes from trials

257 16 to 25 and 26 to 35, but the magnitude of these changes is small (-1.74 and 1.54, respectively).

9

258   For the concussed model, all changes in slope are statistically significant and large in magnitude,

259   ranging from about -15 to 17. The features estimated from the linear spline mixed models are

260   consistent with patterns that are observed in Figure 4.

261   The coefficients from the linear spline mixed model can be combined to estimate the magnitude

262   of the slope in each trial interval. Table 3 presents the estimated slope within each trial range

263   and the corresponding statistical significance using a test of contrasts. Together, results in

264   Tables 2 and 3 suggest differences in average response time patterns for the exposed (hypoxic

265   and concussed) versus non-exposed subjects in both the altitude and concussion data sets. In the

266   two studies, both groups' average response times show a downward slope in the first 5 or so

267   trials, suggesting a learning effect. However, while the normal subjects don't exhibit large

268   changes after that time, the hypoxic and concussed subjects fluctuate throughout the trials, never

269   achieving a constant effect. In the altitude data set, most of the changes in the shape of the

270   average trial profiles are statistically significant for hypoxic subjects and all changes are large

271   and significant for concussed subjects. The significant changes in slopes (i.e. the coefficients of

272   the spline covariates in Table 2) reflect the curvature of the trial profiles for exposed subjects as

273   opposed to the stable trial profiles of the non-exposed subjects following the initial learning

274   effect. Furthermore, the slopes in each trial interval (i.e. the cumulative coefficient effects in

275   Table 3) are large for exposed subjects–ranging from -1.82 to 4.01 and -9.80 to 4.80 for hypoxic

276   and concussed subjects, respectively–after the initial learning effect. In contrast, the slopes in

277   the same intervals are much smaller for the non-exposed: -1.23 to 0.92 and -0.97 to 0.77 for sea

278   level and non-concussed subjects, respectively, reflecting the stability of post-learning responses

279   for these groups.

280   By fitting a combined model for each dataset containing an interaction terms for the exposed vs.

281   non-exposed groups, differences in mean profiles between the two groups were formally tested

282   in each data set (Table 4). In the altitude data set, the change in slope at trial 15 and 25 are

283   statistically significantly different for sea level versus hypoxic subjects, reflecting difference in

284   the mean trial profiles between the groups in the mid-range of the trials. In the concussion data

285   set, the combined model showed that changes in slopes at trial 15, 25 and 35 are significantly

286   different between concussed and non-concussed subjects, indicating that the model identifies

287   different shapes in the average trial profiles between the two groups. For example, for concussed

288   subjects the slope from trial 6 to 10 and trial 16 to 25 is 4.78 (-12.11+16.89) and -5.70 (-

289   12.11+16.89-10.48), respectively. This change in slope, -10.48, is statistically significant. For

290   healthy subjects, the slope from trial 6 to 10 and trial 16 to 25 is 0.76 (-12.11-1.80+16.89-2.22)

291   and -.98 (-12.11-1.80+16.89-2.22-10.48+8.74), respectively. This change in slope, -1.74, is

292   statistically significantly different (and smaller) than the change in slope for the concussed

293   subjects (-10.48).

294   **Discussion**

295   Using simple plots, loess curves, and spline regression—techniques that are readily available in

296   most statistical software packages—we show the potential value of trial-by-trial analysis in

297   discriminating SRT response patterns between concussed and non-concussed subjects and

298   between the same individuals at sea level and extreme altitude. These methods used all available

299   data points and demonstrated significant differences in the shapes of response curves between

300   exposure groups in each data set. Not only were the response curves of exposed subjects more

301   variable than their non-exposed counterparts, but the placement of exposed subjects' curves was

302   also distinct in the two data sets: In the altitude data set, the curve for hypoxic subjects was

303   higher on the Y-axis than that of sea level subjects, while the curve for concussed subjects

304   fluctuated in the same region of the Y-axis as the curve for non-concussed subjects. Thus,

305   although our models demonstrated that the SRT curves for both hypoxic and concussed subjects

306   were highly variable and differed significantly from their non-exposed counterparts, examination

307   of the same data using only means as a strategy to summarize the data would have indicated only

308   that average SRT of hypoxic subjects was higher than at sea level. Thus, although some of the

309   differences between groups can be captured in traditional metrics like means and standard

310   deviations, these summary measures do not capture a wealth of potentially relevant information

311   that is offered by examining the shape of response curves over the full course of test

312   administration. We believe that this information is potentially important and that its value for

313   identifying and tracking cognitive deficit warrants further investigation..

314   While standard t-tests indicated differences of -38.12 (p-value = 0.00) and -4.04 (p-value = .50),

315   for the hypoxic and concussed subjects, respectively, examination of trial-by-trial data yielded a

316   more detailed picture of test performance. Among the sea level and non-concussed subjects,

317   there were stable aggregated mean response times following an initial learning effect. For these

318   subjects, ignoring the time dimension masks the learning effect, but otherwise captured the mean

319   response time. In contrast, the trial-by-trial approach for the hypoxic and concussed subjects

320   demonstrates a curvy response pattern throughout the SRT test administration, and this shape

321   was entirely masked when the data were collapsed across trials and summarized as a mean. In

322   fact, while aggregated analysis indicated no difference in mean response times between groups in

323   the concussion data, the trial-by-trial approach indicated statistically significant differences in

324   mean trial profiles between groups. Thus, our approach to describing cognitive performance has

325   the potential to unmask significant differences between groups that can't be identified with

326   aggregated strategies, thereby allowing examination of research questions that require a more

327   granular approach to the data. A critical question for moving this work forward is the extent to

328   which these methods—which currently focus on group profiles—can be extended to individual-

329   level data.

330   While the same SRT test module was administered to all subjects in this report, important

331   differences in the nature of the altitude and concussion data must be considered. The altitude

332   study was designed to have repeated DANA assessments on the same subjects at various

333   altitudes (i.e. a paired comparison). As a result, the altitude data have a comparable number of

334   subjects for both the sea level and hypoxic conditions. In contrast, the number of subjects is

335   unbalanced in the concussion data set, which includes a relatively small number of concussed

336   subjects from a large sample of cadets. Despite these differences in design and sample size, the

337   trial profiles for the sea level and non-concussed subjects were similar: a dip in average response

338   time until about trial 5, followed by linear and flat average response times for the remaining

339   trials.

340   Importantly, the average trial-level response time was about 300ms among non-exposed subjects

341   in both datasets, reflecting an expected level of similarity among these young, healthy college-

342   age subjects. In contrast, average trial profiles for the concussed and hypoxic subjects were

343   considerably more variable than their non-concussed and sea level counterparts. The relatively

344   larger degree of variability among concussed participants is likely due to small sample size.

345   Nonetheless, although the loess curve for concussed subjects exhibits more curvature than the

346   hypoxic subjects, particularly after trial 10, the knots—the places where participants' slopes

347   change across the 40 trials—were roughly similar in the two groups. Although larger sample

348 sizes are needed to confirm that these methods are effective in distinguishing groups, these
349 empirical observations support the potential utility of focusing on the *pattern* of repeated
350 measures as well as means and standard deviations when assessing differences in SRT in the
351 setting of concussion and hypoxia. These findings also support extension of this line of
352 investigation to other conditions such as depression, PTSD mild cognitive impairment, and
353 dementia.

354 As we develop this line of work, we envision developing the idea of a "cognitive signature" that
355 quantitatively describes an individual's pattern of repeated, computerized cognitive testing
356 measures. The concept of the cognitive signature has been introduced in the setting of bipolar
357 disorder,[26] anorexia nervosa,[27] and multiple sclerosis[28] to describe behavior or symptom patterns
358 that occur with high frequency among patients with specific health conditions. In contrast to
359 these applications, we will describe trial-by-trial cognitive testing data that, with additional
360 investigation and validation, may provide subtle clues to the presence of mild or subclinical
361 cognitive impairment, differences in etiology among various conditions that result in gross
362 cognitive decline, or a more sensitive strategy to tracking cognitive decline over time. The
363 similarity of knot points in our spline regression models for both concussed and hypoxic patients
364 suggests that there may be direct clinical value in pursuing these questions.

365 Importantly, the idea of developing and validating a clinically relevant cognitive signature may
366 not require a full neurocognitive battery. Full batteries are lengthy, labor intensive, and
367 expensive. These characteristics limit their practicality for the "bedside" or "in-clinic" approach
368 needed in the patient care setting, and they add cost to the drug development process. In order to
369 be practical for these varied purposes, a cognitive test should be easy to administer, short in
370 duration, and acceptable to the patient. Ideally, it should also help identify the presence of
371 cognitive deficit, as well as the origin of any observed deficit so that appropriate treatment can
372 be initiated. Advances in technology—particularly mobile technologies—may simplify
373 cognitive assessment, including those that can drive the development of cognitive signatures.
374 These technologies can also facilitate storage and tracking of complex, pattern-rich data over
375 long periods of time, thereby aiding clinicians in early identification of clinically relevant
376 changes in cognitive performance. In addition to their role in screening and tracking change,
377 these strategies could have a significant impact on drug development if distinct cognitive

13

378   signatures are linked to specific health conditions and if their measurement is validated for

379   clinical trials. Indeed, a recent commentary on the value of digital technologies in cognitive

380   assessment focuses on how these technologies provide a "richer, scalable, and objective set of

381   measurements" and these approaches to cognitive testing offer major advantages over the

382   "classically noisy, subjective, data-poor clinical endpoints" that have been used in the past."[29]

383   Further investigation into the role of cognitive signatures in various states of health and disease

384   may help respond to the need for inexpensive methods to detect early features of cognitive

385   decline that leverage the richness of computerized cognitive testing data.

386

387

388

389

390

391

392

Table 1: Summary of altitude and concussion data sets

	Altitude Data	
Study Condition	5260 m	130 m
Number of Subjects	21	17
Average Number of Trials Per Subject	37.4	39.8
Range of Number of Trials Per Subject	10-40	38-40
Average Response Time	337.1	299.0
Range of Response Time	160-863	198-757
	Concussion Data	
Study Condition	Concussed	Not Concussed
Number of Subjects	6	153
Average Number of Trials Per Subject	39.3	39.8
Range of Number of Trials Per Subject	36-40	36-40
Average Response Time	314.8	310.7
Range of Response Time	180-832	181-891

393

394

Table 2: Coefficient Estimates from Linear Spline Mixed Model by Data Set

	Estimate (s.e.)			
	Altitude Data		Concussion Data	
Covariate	Hypoxia	Sea Level	Concussed	Non-Concussed
Intercept	394.4 (26.26)**	344.58 (14.71)**	358.74 (52.41)**	374.71 (6.97)**
Trial Number (Slope: Trials 1-5)	-14.86 (5.72)**	-9.20 (3.08)**	-12.07 (8.27)	-13.90 (1.49)**
Spline: Knot at 5 (Change in Slope @ Trial 5)	18.87 (6.83)**	8.23 (3.68)**	16.87 (9.89)*	14.68 (1.79)**
Spline: Knot at 15 (Change in Slope @ Trial 15)	-7.48 (2.92)**	1.89 (1.59)	-10.50 (4.28)**	-1.74 (0.77)**
Spline: Knot at 25 (Change in Slope @ Trial 25)	6.47 (2.86)**	-0.34 (1.58)	10.40 (4.27)**	1.54 (0.76)**
Spline: Knot at 35 (Change in Slope @ Trial 35)	-4.82 (5.40)	-1.81 (2.98)	-14.51 (8.01)*	-0.74 (1.44)

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively

395

Table 3: Slopes of Linear Spline Mixed Model for Trial Number Intervals

Trial Number Range	Estimate (s.e.)			
	Altitude Data		Concussion Data	
	Above Sea Level	Sea Level	Concussed	Healthy
1-5	-14.86 (5.72)**	-9.20 (3.08)**	-12.07 (8.27)	-13.91 (1.49)**
6-15	4.01 (1.71)*	-0.97 (0.93)	4.80 (2.50)	0.77 (0.45)
16-25	-3.47 (1.54)*	0.92 (0.84)	-5.70 (2.27)*	-0.97 (0.41)*
26-35	3.01 (1.65)	0.58 (0.91)	4.71 (2.47)	0.57 (0.44)
>35	-1.82 (4.28)	-1.23 (2.34)	-9.80 (6.32)	-0.17 (1.14)

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively

396

397

398

399

Table 4: Linear Spline Mixed Model Coefficient Estimates – Combined Model

Covariate	Estimate (s.e.)	
	Altitude Data	Concussion Data
Intercept	394.4 (21.45)**	358.80 (35.74)**
Non-Exposed	-49.78 (31.40)	15.91 (36.44)
Trial Number (Slope: Trials 1-5)	-14.86 (4.64)**	-12.11 (7.54)
Non-Exposed * Trial Number	5.65 (6.73)	-1.80 (7.69)
Spline: Knot at 5 (Change in Slope @ Trial 5)	18.87 (5.54)**	16.89 (9.02)*
Spline: Knot at 15 (Change in Slope @ Trial 15)	-7.48 (2.37)**	-10.48 (3.91)**
Spline: Knot at 25 (Change in Slope @ Trial 25)	6.47 (2.32)**	10.43 (3.89)**
Spline: Knot at 35 (Change in Slope @ Trial 35)	-4.81 (4.38)	-14.56 (7.31)**
Non-Exposed* Spline: Knot at 5	-10.64 (8.05)	-2.22 (9.20)
Non-Exposed * Spline: Knot at 15	9.37 (3.46)**	8.74 (3.98)**
Non-Exposed * Spline: Knot at 25	-6.81 (3.41)**	-8.89 (3.97)**
Non-Exposed * Spline: Knot at 35	3.00 (6.45)	13.82 (7.45)*

Note: Statistical significance at the 95% and 90% confidence level is indicated by ** and *, respectively.

400

401

402

403 Figure 1: Trial Profiles (solid lines) and Trial Averages (points) of Response Time by Exposure
404 in the Altitude Data Set
405

406

407
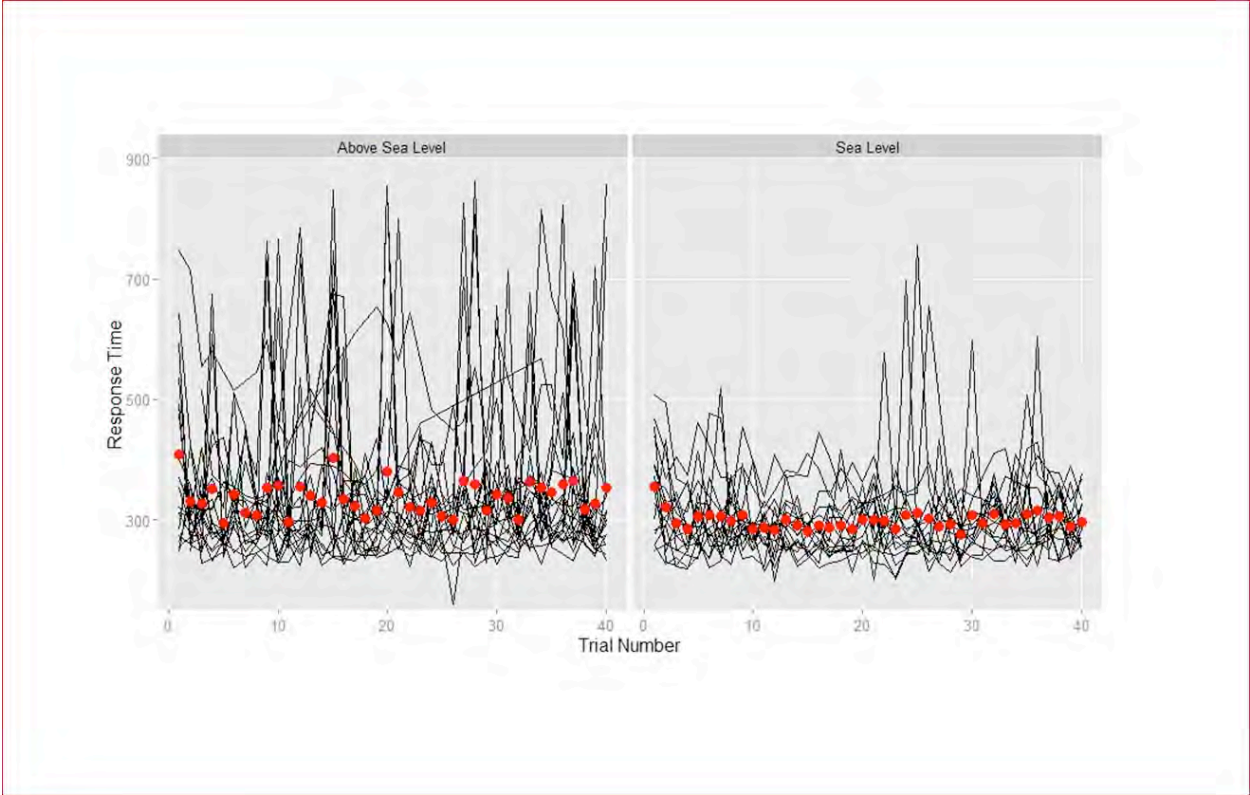
408   Figure 2: Trial Profiles (solid lines) and Trial Averages (points) of Response Time by Exposure
409   the Air Force Data Set
410

411

412
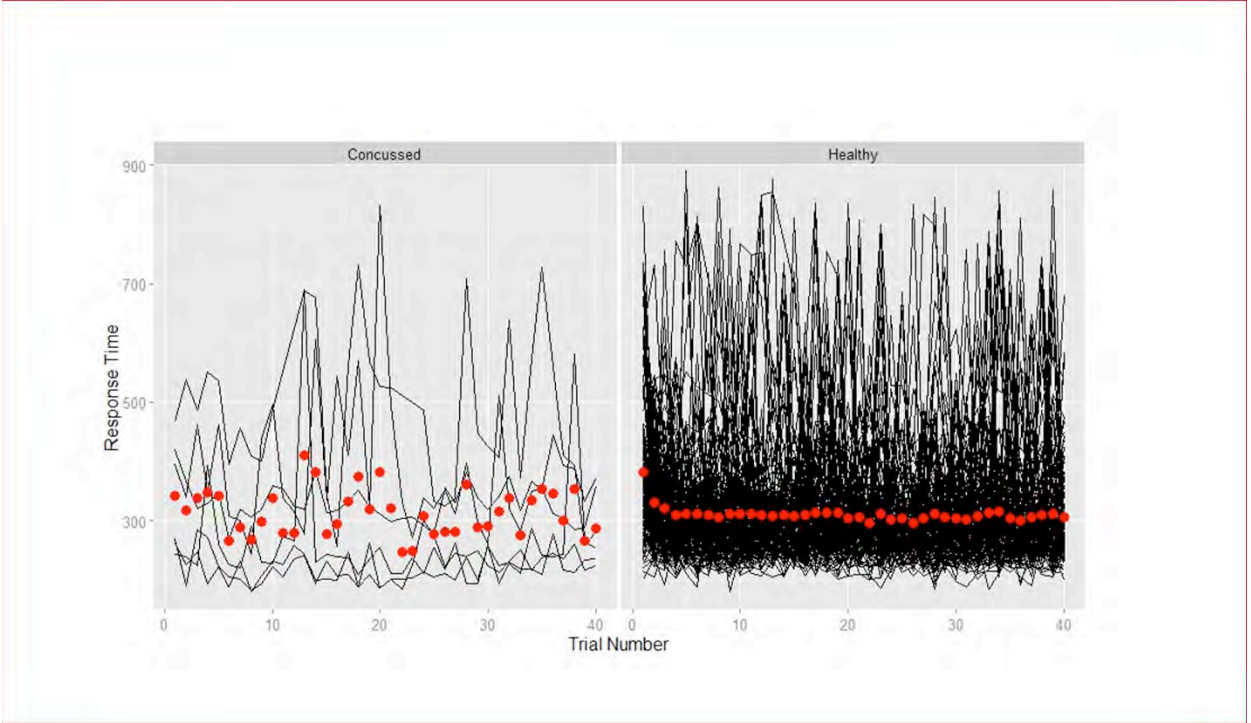
413 Figure 3: Trial Averages (points), Loess (dashed line) and Linear Spline Regression Fit (solid
414 line) of Response Time by Elevation in the Altitude Data Set
415

416

417
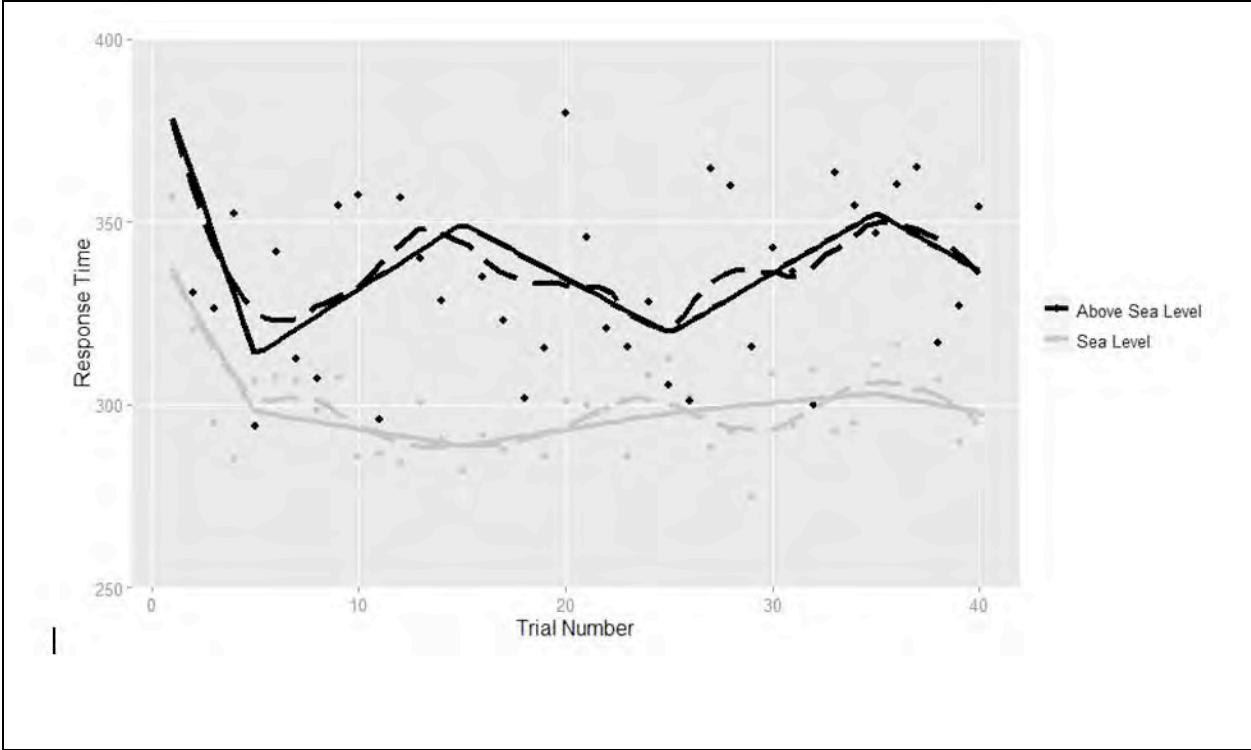
418 Figure 4: Trial Averages (points), Loess (dashed line) and Linear Spline Regression Fit (solid
419 line) of Response Time by Concussion Status in the Concussion Data Set
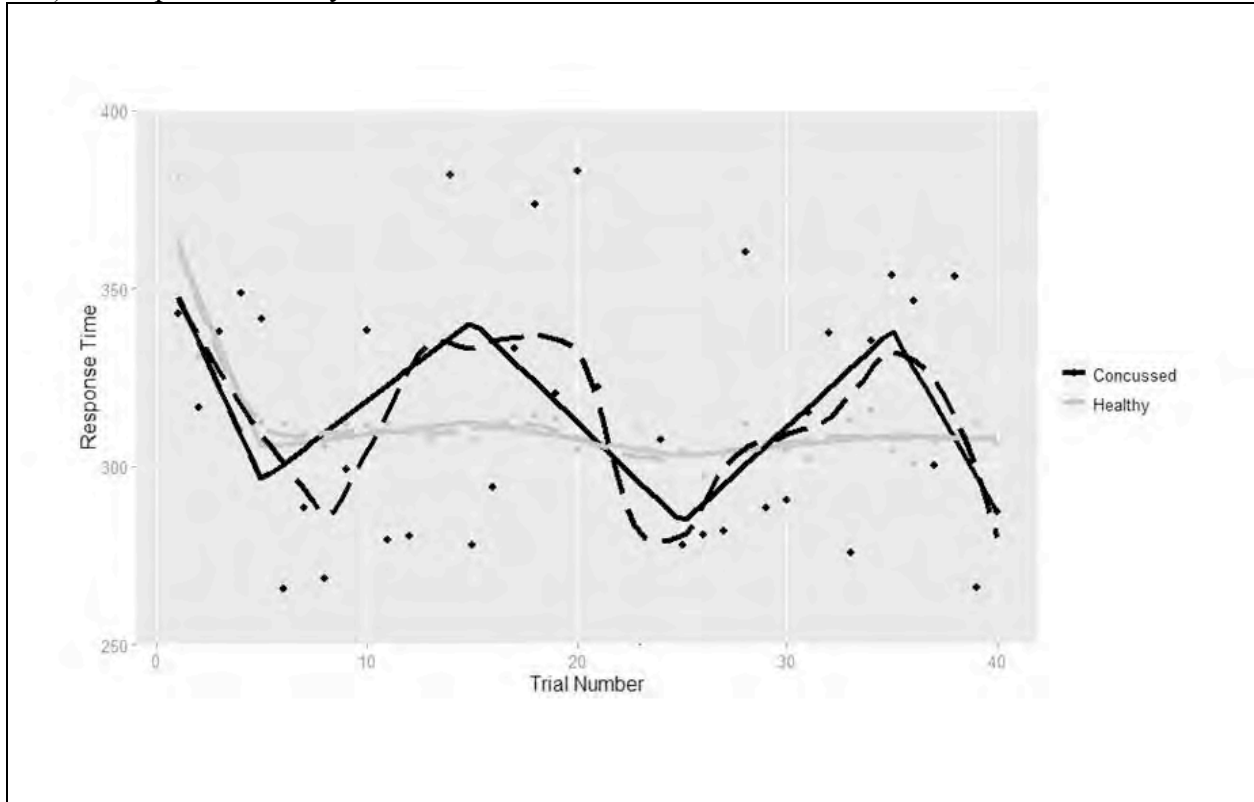
References

[1] Cullen et al. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* (2007) **78**:790–799.

[2] Borson et al. Improving dementia care: The role of screening and detection of cognitive impairment. *Alzheimers Dement* (2013) **9**: 151–159.

[3] http://www.alz.org/health-care-professionals/cognitive-tests-patient-assessment.asp (accessed January 5, 2015)

[4] Roiser, J.P. et al. Assessment of cognitive safety in clinical drug development, *Drug Discov Today.* 2015; http://dx.doi.org/10.1016/j.drudis.2015.11.003.

[5] Wesnes KA and Edgar CJ: The role of human cognitive neuroscience in drug discovery for the dementias. *Curr Opin Pharmacol* (2014) **14**:62–73.

[6] Folstein et al. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* (1975) **12**:189-98.

[7] Avila ML et al: A critical review of scoring options for clinical measurement tools. *BMC Res Notes* (2015) **8**: 612.

[8] Wild et al. The status of computerized cognitive testing in aging: A systematic review. *Alzheimers Dement* (2008) **4**: 428–437.

[9] Wang L et al. Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behav Res* (2009) **43**: 476–496.

[10] Der G and Deary IJ: Age and Sex Differences in Reaction Time in Adulthood: Results From the United Kingdom Health and Lifestyle Survey. *Psychol Aging* (2006) **21:** 62–73.

[11] Snodgrass et al. Some experiments on simple and choice reaction time. *J Exper Psychol* (1967) **75**:1-17.

[12] Collardeau et al. Effects of a prolonged run on simple reaction time of well trained runners. *Percept Mot Skills*. (2001) **93**:679-89.

[13] Hagger-Johnson et al. Reaction Time and Mortality from the Major Causes of Death: The NHANES-III Study. *PLoS one* (2014) **9:** e82959

[14] Lathan, Corinna, et al. "Defense Automated Neurobehavioral Assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool." *Mil Med* (2013) **178**: 365-371.

[15] Russo, C. R., and C. E. Lathan. "An Evaluation of the Consistency and Reliability of the Defense Automated Neurocognitive Assessment Tool." *Appl Psychol Measur* (2015) doi: 0146621615577361.

[16] Spira JL, Lathan CE et al. The impact of multiple concussions on emotional distress, post-concussive symptoms, and neurocognitive functioning in active duty United States Marines independent of combat exposure or emotional distress. *J Neurotrauma* (2014) **31**:1823-34.

[17] Roach, Emma B., et al. "AltitudeOmics: Decreased reaction time after high altitude cognitive testing is a sensitive metric of hypoxic impairment."*NeuroReport* (2014) **25**: 814.

[18] Phillips M, Rogers P, Haworth J, Bayer A, Tales A (2013) Intra-Individual Reaction Time Variability in Mild Cognitive Impairment and Alzheimer's Disease: Gender, Processing Load and Speed Factors. *PLoS ONE.* 8(6): e65712. doi:10.1371/journal.pone.0065712.

[19] Baddeley AD et al. Attentional control in Alzheimer's disease. *Brain* (2001) **124**:1492-1508.

[20] Jordan N et al. Cognitive components of reaction time in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. (1992) **55**: 658–664.

[21] White DA et al. How Cognitive is Psychomotor Slowing in Depression? Evidence from a Meta-Analysis. *Aging, Neuropsychol Cog* (1997) **4:** 166-174.

[22] Edinger JD et al. Psychomotor Performance Deficits and Their Relation to Prior Nights' Sleep Among Individuals with Primary Insomnia. *Sleep* (2008) **31**: 599–607.

[23] Subudhi, Andrew W., et al. "AltitudeOmics: the integrative physiology of human acclimatization to hypobaric hypoxia and its retention upon reascent." *PloS one* (2014) **9**.3: e92191.

[24] Jacoby WG: Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies* (2000) **19:** 577–613.

[25] Wold S. Spline functions in data analysis. *Technometrics* (1974) **16**:1-11.

[26] Glahn DC et al. The Neurocognitive Signature of Psychotic Bipolar Disorder. *Biol Psychiatr* (2007) **62**: 910–916.

[27] Roberts ME et al. Exploring the neurocognitive signature of poor set-shifting in anorexia and bulimia nervosa. *J Psychiatr Res* (2010) **44**: 964–970.

[28] Hanken K et al. Is there a cognitive signature for MS-related fatigue? *Mult Scler* (2015) **21**:376-81.

[29] Leurent C and Ehlers MD. Digital technologies for cognitive assessment to accelerate drug development in Alzheimer's disease. *Adv Neurodegen Dis Demen* (2015) **98**: 475-76.

476
477

Computerized cognitive testing norms in active-duty military personnel:

Potential for contamination by psychologically unhealthy individuals

Ian Coffman, Helaine E. Resnick, James Drane, Corinna E. Lathan

AnthroTronix, Inc.

Abstract

Normative reference data used for clinical interpretation of neuropsychological testing results are only valid to the extent that the sample they are based on is composed of "normal" individuals. Accordingly, efforts are made to exclude individuals with histories and/or diagnoses that might bias test performance. In this report we focus on these features in active-duty military personnel because published data on computerized neurocognitive testing norms for this population have not explicitly considered the consequences of neurobehavioral disorders (e.g., PTSD, depression, etc.), which are prevalent in this population and known to affect performance on some cognitive assessments. We administered DANA, a mobile, neurocognitive assessment tool, to a large sample of active-duty military personnel and found that scores on self-administered psychological assessments negatively impacted a number of neurocognitive tests. These results suggest that neurobehavioral disorders that are relatively common in this population should be controlled for when establishing normative datasets for neurocognitive outcomes.

Keywords: Cognitive assessment, normative data, PTSD, active-duty military

INTRODUCTION

Normative data are used extensively in the clinical interpretation of neurocognitive testing results since a single score on an assessment is difficult to interpret without knowing how it compares to others in the test-taking population. For example, an individual's performance might be compared to a reference distribution centered on a sample mean that is derived from a set of "normal" subjects. Because there is variability in these results–even among normally functioning subjects–cutoff points at each end of the distribution are established as thresholds for classifying patients as potentially "non-normal." If the reference empirical distribution is approximately symmetric, then any patient falling more than +/- 2 standard deviations away from the mean may be considered "non-normal."

It is widely recognized that normative data are only useful to the extent that (i) they can be applied to individuals with similar characteristics to the sample from which the data were collected (e.g., Heaton et al., 1986; Ross & Lichtenberg, 1998) and (ii) that the reference data comprise observations from "normal," i.e., unimpaired, subjects. Appreciation of the latter point has resulted in efforts to exclude from normative data individuals with histories and/or diagnoses that can be reasonably expected to affect test performance, and by extension, the larger distribution. Exclusion criteria can include general, domain-specific features (e.g., history of memory complaints if the normative data are for scores on a memory-based test) as well as features that might be expected to occur with greater incidence in a particular population (e.g., presence of dementia in a geriatric sample). For example, Schneider and colleagues (2015) pre-screened subjects for history of dementia and other age-related neurologic issues when they collected normative data for a number of neuropsychological tests from a sample of older adults.

Demographic characteristics of the sample, e.g., age and gender, may also affect normative data. If the population that the sample is meant to represent includes multiple levels or values of these characteristics (e.g., both males and females for gender) and if differences in these values are thought to yield substantial effects on the measure of interest, then these effects must also be controlled for. Typically, this is accomplished by presentation of stratified tables of means conditioned on each value of the demographic feature. For example, after statistical testing revealed significant effects of age and education, Ganguli et al. (2010) presented normative data for a number of cognitive assessments stratified by both of these features.

In this report, we consider the task of establishing a normative dataset for a neurocognitive assessment tool (NCAT) in the context of the active-duty military population. In particular, we focus on which population-specific features should be accounted for in the process of defining a normative dataset. Beyond controlling for basic demographic factors such as age and gender, most published normative data on NCATs for military use cases (e.g., Vincent et al., 2012; Roebuck-Spencer et al., 2013) do not consider the consequences of neurobehavioral disorders that likely affect active duty service members at a rate greater than that of the general population (e.g., posttraumatic stress disorder (PTSD)). Although Roebuck-Spencer and colleagues identify this issue as a potential limitation of their approach, they note that because their sample "comprised active duty service members who had not been medically discharged, they are presumed to be healthy and free of medical or psychiatric conditions that would significantly impair performance on neuropsychological testing" (p. 503). This report explores this basic assumption.

Despite the implications of service members' discharge status or classification as "active-duty," there is evidence that a subset of this population may be affected by neurobehavioral

challenges. For example, Hoge et al. (2004) estimated the rate of probable PTSD at 9 percent among pre-deployed service members, and 12 and 18 percent at post-deployment among participants in Operations Enduring Freedom and Iraqi Freedom, respectively. Those findings are relevant because PTSD is known to negatively impact performance on certain neurocognitive tests (e.g., Horner & Hamner, 2002; Swick et al., 2012). These results suggest that it is prudent to explicitly test for the presence of various psychological disorders to determine whether their neurocognitive consequences, if any, are extreme enough to substantially bias data that would otherwise be mistakenly classified as "normative."

Using a large sample of active duty service members aged 18-64, we examined performance on eight cognitive tests, and then studied the impact of self-reported sleep disturbance, depression, and PTSD on these distributions. We hypothesized that independent of age and gender, active duty military personnel with sleep disturbances, depression, and PTSD would perform less favorably on computerized cognitive tests than their counterparts who do not report these conditions. If this hypothesis proves true, it has direct implications for how normative data are used to evaluate cognitive efficiency in active duty military personnel.

METHODS

DANA, a hand-held, computerized neurocognitive assessment tool, was administered to 814 active duty service members (71% male) aged 18-64 stationed at the Fort Hood military post near Killeen, Texas. A data collection error resulted in three instances in which two participants' data were assigned to a single unique identifier. These six records were excluded from analysis, yielding a final sample size of 808.

Service members were recruited via distribution of fliers at locations around the Ft. Hood post and through direct briefings by the site PI after unit/company formation. Consent materials stated that the research goal was to collect a large database of cognitive data on active-duty military, so participants were naïve to any potential theoretical comparisons. It should be noted that since 59 percent of our sample had been previously deployed, depending on when deployment occurred, some may have been exposed to the Automated Neuropsychological Assessment Metrics (ANAM) battery (e.g., Vincent et al., 2012). A 2008 Congressional mandate required administration of this battery prior to deployment. ANAM's subtests are comparable to DANA's, so it is possible that some participants entered the experimental setting having prior experience with computerized cognitive testing.

Military personnel were eligible to take part in the study if they were classified as active duty and between the ages of 18 and 64 (inclusive). Potential participants were excluded if they had consumed alcohol within the last eight hours, regularly used mind-altering medications (e.g., anti-psychotic medications, benzodiazepines, Benadryl, etc.), or had sustained a concussion within the month prior to testing. DANA was administered on Samsung Galaxy S4 smartphones, which the developers of DANA have found to be technically suitable for this purpose.

DANA contains a battery of tests designed to examine cognitive performance on several tasks, and it also includes several psychological tests. Its favorable psychometric properties and test-retest reliability has been documented (Lathan et al., 2013; Russo & Lathan, 2015). Russo & Lathan demonstrate that the test-retest reliability coefficients for DANA's Simple Reaction Time and Procedural Reaction Time subtests (intraclass correlation coefficients of 0.81 and 0.75, respectively) are comparable to other assessment batteries that contain these tests. A summary of the neurocognitive tests examined in this study is provided in Table 1.

A score of less than 66 percent correct on any DANA subtest is considered an invalid administration and excluded from analysis. This criterion is evaluated on a per-subtest basis; if a participant scored less than 66 percent correct on a given subtest or set of subtests, only those observations are excluded, but the remainder of their record is included in analysis. The main outcome variable in this study is "throughput," a speed-accuracy product that quantifies the number of correct responses per minute (Thorne, 2006):

$$\text{Accuracy} \times \text{Speed} \times 60{,}000$$

where accuracy is the proportion of correct responses, speed is the reciprocal of mean correct reaction time, and the scaling factor of 60,000 converts the quantity to units of min^{-1}.

Participants were also administered three psychological assessments: the Posttraumatic Stress Disorder Checklist - Military (PCL-M), a military-specific posttraumatic stress disorder assessment (McDonald & Calhoun, 2010), the Patient Health Questionnaire 8 (PHQ-8), a depression diagnostic (Kroenke et al., 2009), and the Pittsburgh Sleep Quality Index (PSQI), a measure of sleep disturbance/insomnia (Buysse et al., 1989).

Analysis

The goal of this report is to examine the effects of PCL-M (PTSD), PHQ-8 (depression) and PSQI (insomnia) scores on cognitive performance, and to understand the impact of these measures on normative data in active duty military personnel.

We assess relationships between throughput and scores on the three psychological assessments with regression models that control for age and gender. Given this strategy, an

immediate issue to consider is the expected pattern of comorbidities among the disorders that

these measures assess. For example, the relationships between PTSD and depression (e.g.,

O'Donnell et al., 2004; Brady et al., 2000) and PTSD and disturbed sleep (e.g., Mysliwiec et al.,

2013; Leskin et al., 2013) have received considerable attention in the literature. If present in our

sample, evidence of these associations would have important implications for both the

conceptual and data-analytic components of our study. Accordingly, we first examined

correlations among PCL-M, PHQ-8 and PSQI scores to inform our general strategy for modeling

the data.

Statistical analyses were carried out under R version 3.3.1 (R Core Team, 2016).

Regression models were fit via ordinary least squares, and visual inspection of fitted vs. residual

and normal quantile-quantile plots indicated that model assumptions were adequately met, and

consideration of Cook's distance revealed that no data points had undue influence on parameter

estimates. Gender and age were considered potential confounders and included as covariates, and

they were treated as categorical and dummy coded. For age, the 18-19 band serves as the

reference level, and the female gender category serves as the reference for gender. Reported

coefficients are unstandardized so that effect sizes can be interpreted in terms of throughput, the

unit of interest.

RESULTS

Table 2 provides a breakdown with marginal totals of the final sample age-gender distribution.[1]

---

[1] These age groups are hard-coded into the demographic questionnaire included in DANA and were thus not devised with reference to the present analysis.

DANA throughput values are approximately normally distributed, but this is not true of scores on the administered psychological assessments. Table 3 provides descriptive statistics for these outcomes. The distributions of scores for all three assessments are highly right-skewed, with extreme scores on the high end of the distribution suggesting outliers. We take this asymmetry to reflect the general incidence of the disorders they measure and note that including all scores in the regression analyses described below yielded models that display an adequate linear fit to the data, suggesting that while relatively few in number, these extreme scores are nonetheless principled.

As Table 4 indicates, scores on these assessments are highly correlated. Given these correlations, multicollinearity is likely to prevent isolation of the unique effect of each psychological assessment score on throughput outcomes. We therefore employed a hierarchical procedure in which two regression models were fit for each DANA subtest. The first regresses throughput on the control variables (age and gender) and PCL-M scores.[2] The second is identical but was further specified to include covariates for PHQ-8 and PSQI scores. We used the first model to examine the coefficient associated with PCL-M scores, and the second was compared to the first, allowing assessment of whether PHQ-8 and PSQI scores significantly impact throughput beyond the effect of PCL-M scores alone.[3]

Results of the PCL-M-only models reveal a significant negative effect of PCL-M scores for five of the eight DANA subtests: SRT1, PRT, GNG, CSL and SRT2 (Table 5). Consistent with findings from other normative studies of neurocognitive assessment (e.g., Vincent et al.,

---

[2] We chose to include PCL-M scores as a predictor in our reduced models under the assumption that it would explain the most variance in the data relative to PHQ-8 and PSQI scores. This assumption is based on the implicational relationships present among the three disorders these measures assess: in terms of symptoms, PTSD can imply depression and sleep issues, but the reverse is not necessarily true. PCL-M scores were scaled from their original range (17-85) to 0-68 to facilitate interpretation of the intercept term.
[3] We adopted this strategy because multicollinearity issues are only problematic for inference on individual coefficients within a model and not inference on model comparison as summarized by an ANOVA F-ratio.

2012), significant effects of age and gender are also observed in the expected direction for a number of subtests as well.[4] We also point out that our validity criterion, i.e., greater than 66 percent of trials correctly completed, results in different numbers of exclusions depending on the subtest. A particular issue concerns the Code Substitution – Recall (CSR) subtest, where 13 percent of the observations were excluded as invalid. This is likely due to this subtest's greater difficulty relative to others.

Figure 1 plots age- and gender-adjusted PCL-M slopes to facilitate a comparison of effect sizes across the DANA subtests where the PCL-M coefficient reached significance.

For each subtest, the PCL-M-only model was compared to a model including covariates for PHQ-8 and PSQI scores via ANOVA. The results of these analyses show that inclusion of PHQ-8 and PSQI scores provides very little additional explanatory power over inclusion of the PCL-M covariates alone (Table 6), suggesting that PCL-M accounts for a large majority of the variance among psychological assessment covariates.

DISCUSSION

This report focused on the task of establishing a normative dataset of neurocognitive performance for the active-duty military population. This work was driven in part by previous research describing the incidence and consequences of behavioral issues in active-duty military samples. For example, Spira et al. (2014) documented negative relationships between PCL-M and PHQ-8 scores and concussion history and neurocognitive performance, a result that suggests behavioral issues may impact neurocognitive abilities in this population in a more direct fashion.

---

[4] Treating PCL-M scores as a continuous variable renders the presentation stratified normative tables impossible. However, an individual's mean throughput values can be predicted from the regression equations. For example, the expected SRT1 throughput value for a 32 year-old male with a PCL-M score of 30 can be calculated as follows: $182.34 - 0.38*(30-17) + 9.42 - 3.69 = 183.13$. See Van Breukelen & Vlaeyen (2005) for more detail on the regression-based approach to normative data.

The potential implications of these questions can be appreciated in the context of results from a meta-analysis of 25 studies that estimated the prevalence of DSM-IV major depression among U.S. military personnel. That study showed that the prevalence of depression was 12.0% among current deployed personnel and 13.1% among previously-deployed personnel (Gaderman et al., 2012). Behavioral issues such as depression have a causal impact on cognitive function (e.g., slowed reaction time; Azorin et al., 1995), therefore the relatively high prevalence of these psychological risk factors suggests that they may play a potentially large role in influencing the distribution of "normative" cognitive function measures in active-duty military personnel.

Our data showed that scores on the PCL-M, an instrument that assesses PTSD severity, were also negatively associated with performance on a number of neurocognitive tests in the DANA battery. This basic observation extends the implications of our findings concerning establishment of normative data in the active duty military population. We also observed strong correlations among PCL-M, PHQ-8, and PSQI scores. These correlations reflect an expected comorbidity pattern among these factors and suggests that these factors tend to cluster among affected individuals. We also found that PHQ-8 and PSQI scores did not account for a significant portion of the variance in throughput beyond what is contributed by PCL-M scores alone. Thus, although PTSD severity, sleep disturbance, and depression are associated with one another, in our sample, the impact of these factors on neurocognitive performance can be explained adequately using only PCL-M.

An immediate application of these findings concerns the establishment of normative datasets for neurocognitive performance among active-duty military personnel. The consequences of neurobehavioral disorders on testing results have not been adequately examined in this setting. We suspect this gap in the literature is due in part to the assumption that an

"active-duty" designation implies a population free of psychiatric disorders that would affect performance on neurocognitive assessments, as suggested by Roebuck-Spencer et al., 2013. To our knowledge, this is the first study to this issue. Our findings challenge the basic assumption that an active-duty designation is sufficient to define the population from which normative data can be derived for military personnel.

In particular, our models show that increasing PCL-M scores cause the mean of the throughput distribution to decrease. If the effect of these scores is not controlled for, then the false negative rate for psychologically healthy individuals will increase. This is because any "non-normal" threshold, e.g., greater than two standard deviations below the mean, etc., itself depends on the location of the distribution. If the mean is downwardly biased, then more individuals who are "non-normal" in the unbiased distribution will be classified as "normal" when the biased distribution is utilized for comparison. The extent of misclassification due to this bias depends on its effect size, with larger effects resulting in a greater number of misclassifications.

Extending beyond the issue of what population should be used to define normative neuropsychological data among active-duty military personnel, results of this study have implications for a more basic issue concerning the utility of normative neuropsychological data: specifically, the fundamental challenge of knowing which features of a population to measure and control for to ensure that a normative data set truly represents the performance of normally functioning individuals. While it is possible to identify many features by examining both domain- and population-specific factors that might reasonably be expected to affect cognitive performance, others will invariably be missed.

A final issue concerns the interpretation of our results in terms of what constitutes a "normal" reference sample. If the incidence of PTSD and related comorbidities is relatively high in a population, then can a sample that includes affected individuals be considered "normal?" If the complete sample were utilized, it would be necessary to control for the effects of these disorders (e.g., via regression-based norms with appropriate covariates or via stratified tables of conditional means). On the other hand, it can be argued that it would be appropriate to exclude these individuals if their psychological features are thought to be generally uncharacteristic of the target population under consideration.

In our data, 13 percent of participants scored within clinical range on the PCL-M (≥ 34 for "moderate PTS"), six percent within clinical range on the PHQ-8 (≥ 10 for "major depression") and 48 percent within clinical range on the PSQI (≥ 5 for "poor sleep quality"). If the definition "normal" for this population is based in part on these frequencies, then it might be difficult to justify excluding individuals with evidence of poor sleep quality since they comprise nearly half of the sample. On the other hand, relatively few participants scored within the clinical range for depression, so they might be considered "outside the norm" and excluded from the sample with little loss of power related to the reduction in sample size. Although the purpose of this report was not to argue a position on this issue, we highlight this particular practical consequence of our findings, which reflects the original purpose for pursuing this line of investigation.

Two important limitations of this study should be addressed. First, evidence of neurobehavioral issues (PTSD, depression, and disturbed sleep) were obtained via self-report, and no formal diagnoses were obtained. Although the instruments used to assess these disorders have been validated, they are not perfectly predictive of clinical diagnosis. However, if the

scores obtained on these assessments are in some cases not truly reflective of the underlying construct they are meant to measure, this issue only affects the interpretation of our results and not their utility. The absence of clinical diagnosis does not alter the fact that participants' self-assessments correlate negatively with neurocognitive performance; only the link between these measures and the construct they assess is under question.

A related issue is that we considered only a limited number of neurobehavioral issues and a limited number of instruments for their assessment. It is possible that other assessments, or a combination of assessments, would provide a more accurate diagnostic of psychological disorder. Further, there are other disorders we did not assess, such as anxiety, that may also negatively impact neurocognitive testing results. We expect that future research efforts will consider this issue in more detail.

CONCLUSION

This report examined key underlying assumptions concerning the use of normative data for cognitive testing in active-duty military personnel. We show that scores on assessments for PTSD, depression, and disturbed sleep —psychological issues that occur with relatively high frequency among active-duty personnel—have an unfavorable impact on quantitative measures of cognitive efficiency. These psychological factors may therefore skew the distributions of cognitive efficiency measures in large samples of seemingly healthy military personnel in ways that could affect "non-normal" classifications.

ACKNOWLEDGEMENTS

DISCLOSURE STATEMENT

AnthroTronix, Inc., is the developer of DANA.

REFERENCES

Azorin, J. M., Benhaim, P., Hasbroucq, T., & Possamaï, C. A. (1995). Stimulus preprocessing

and response selection in depression: A reaction time study. *Acta Psychologica*, *89*(2),

95-100.

Brady, K. T., Killeen, T. K., Brewerton, T., & Lucerini, S. (2000). Comorbidity of psychiatric

disorders and posttraumatic stress disorder. *Journal of Clinical Psychiatry*, *61*, 22-32.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The

Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.

Psychiatry Research, *28*(2), 193-213.

Gadermann, A.M., Engel, C.C., Naifeh, J.A., Nock, M.K., Petukhova M., Santiago, P.N.,Wu, B.,

Zaslavsky A.M., Kessler, R.C. (2012) Prevalence of DSM-IV major depression among

U.S. military personnel: Meta-analysis and simulation. Military Medicine. 177(80): 47–

59.

Ganguli, M., Snitz, B. E., Lee, C. W., Vanderbilt, J., Saxton, J. A., & Chang, C. C. H. (2010).

Age and education effects and norms on a cognitive test battery from a population-based

cohort: The Monongahela–Youghiogheny Healthy Aging Team. *Aging and Mental

Health*, *14*(1), 100-107.

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, *351*(1), 13-22.

Horner, M. D., & Hamner, M. B. (2002). Neurocognitive functioning in posttraumatic stress disorder. *Neuropsychology review*, *12*(1), 15-30.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1), 163-173.

Lathan, C., Spira, J. L., Bleiberg, J., Vice, J., & Tsao, J. W. (2013). Defense Automated Neurobehavioral Assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool. *Military Medicine*, *178*(4), 365-371.

Leskin, G. A., Woodward, S. H., Young, H. E., & Sheikh, J. I. (2002). Effects of comorbid diagnoses on sleep disturbance in PTSD. *Journal of Psychiatric Research*, *36*(6), 449-452.

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: a critical review. *Clinical Psychology Review*, *30*(8), 976-987.

Mysliwiec, V., McGraw, L., Pierce, R., Smith, P., Trapp, B., & Roth, B. J. (2013). Sleep disorders and associated medical comorbidities in active duty military personnel. *Sleep*, *36*(2), 167-74.

O'Donnell, M. L., Creamer, M., & Pattison, P. (2004). Posttraumatic stress disorder and depression following trauma: understanding comorbidity. *American Journal of Psychiatry*, *161*(8), 1390-1396.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation

      for Statistical Computing, Vienna, Austria. URL: https://R-project.org.

Roebuck-Spencer, T. M., Vincent, A. S., Schlegel, R. E., & Gilliland, K. (2013). Evidence for

      added value of baseline testing in computer-based cognitive assessment. *Journal of*

      Athletic Training, *48*(4), 499.

Russo, C. R., & Lathan, C. E. (2015). An evaluation of the consistency and reliability of the

      defense automated neurocognitive assessment tool. *Applied Psychological Measurement*,

      0146621615577361.

Schneider, A. L., Sharrett, A. R., Gottesman, R. F., Coresh, J., Coker, L., Wruck, L., ... &

      Mosley, T. H. (2015). Normative Data for Eight Neuropsychological Tests in Older

      Blacks and Whites from the Atherosclerosis Risk in Communities (ARIC) Study.

      Alzheimer Disease and Associated Disorders, *29*(1), 32.

Spira, J.L., Lathan C.E., Bleiber J., Tsao J.W. (2014). The Impact of Multiple Concussions on

      Emotional Distress, Post-Concussive Symptoms, and Neurocognitive Functioning

      in Active Duty United States Marines Independent of Combat Exposure or Emotional

      Distress. *Journal of Neurotrauma,* 31:1823–1834.

Swick, D., Honzel, N., Larsen, J., Ashley, V., & Justus, T. (2012). Impaired response inhibition

      in veterans with post-traumatic stress disorder and mild traumatic brain injury. *Journal*

      of the International Neuropsychological Society, *18*(05), 917-926.

Thorne, D. R. (2006). Throughput: a simple performance index with desirable characteristics.

      Behavior Research Methods, *38*(4), 569-573.

Van Breukelen, G. J., & Vlaeyen, J. W. (2005). Norming clinical questionnaires with multiple

      regression: the Pain Cognition List. *Psychological Assessment*, *17*(3), 336.

Vincent, A. S., Roebuck-Spencer, T., Gilliland, K., & Schlegel, R. (2012). Automated

    Neuropsychological Assessment Metrics (v4) traumatic brain injury battery: military

    normative data. *Military Medicine*, *177*(3), 256-269.

Table 1: Description of DANA Subtests

Test name	Task Description
Simple Reaction Time (SRT)	The subject taps an orange target symbol as quickly as possible each time it appears. The location and shape of the stimulus does not vary from trial to trial.
Procedural Reaction Time (PRT)	The screen displays one of four numbers (1, 2, 3, or 4) for 2 seconds. The subject taps the left button ("2 or 3") or right button ("3 or 4") as quickly as possible to indicate which category

	corresponds to the number displayed.
Go/No-Go (GNG)	A building is presented on the screen with several windows. Either a "friend" (green) or "foe" (gray) appears in a window. The subject must tap the "BLAST" button as quickly as possible only when a "foe" appears.
Code Substitution – Learning (CSL)	Subjects refer to a key of 9 symbol-digit pairs that are shown across the upper portion of the screen. Single symbol-digit pairs are presented in succession below the key, and the subject indicates whether or not the single pair matches the code by tapping "Yes" or "No" As quickly as possible.
Code Substitution – Recall (CSR)	After a delay of several intervening tests, the same symbol-digit pairs from the earlier Code Substitution – Learning task are presented without the key. The subject indicates whether or not the pairing was included in the code that was presented in the earlier Code Substitution – Learning section by tapping "Yes" or "No" as quickly as possible.

Spatial Processing (SP)	Pairs of four-bar histograms are displayed on the screen, one pair at a time and simultaneously, with one histogram rotated 90 degrees (either clockwise or counterclockwise). The subject is required to determine whether the two histograms would be identical if no rotation was applied by tapping either the "Same" or "Different" button as quickly as possible.
Matching to Sample (MTS)	A single 4 x 4 checkerboard pattern is presented on the screen for a study period of 3000 ms. It then disappears for 5 seconds, after which two patterns are presented side-by-side. The subject indicates which of these two patterns was displayed during the study period by tapping on the checkerboard that they believe is identical to the originally presented stimulus as quickly as possible.

Table 2: Sample age and gender distribution

	18-19	20-24	25-29	30-34	35-44	45-54	55-64	Total
Male	99	100	100	94	91	58	29	571

Female	24	94	51	29	23	13	3	237
Total	123	194	151	123	114	71	32	**808**

Table 3: Descriptive Statistics for Administered Psychological Assessments

Assessment	Possible range	Min	1st quartile	Median	3rd quartile	Max	Mean
PCL-M	17-85	17.00	17.00	20.00	26.00	85.00	24.07
PHQ-8	0-24	0.00	0.00	2.00	4.00	22.00	2.80
PSQI	0-21	0.00	2.00	4.00	7.00	16.00	4.79

Table 4: Pairwise Correlations Between Psychological Assessments

	PCL-M	PHQ-8
PCL-M		
PHQ-8	.74***	
PSQI	.66***	.66***

Note: *** = p < .001

Table 5: Coefficient Estimates and (standard errors) by DANA Subtest

β	**SRT1** (N=804)	**PRT** (N=805)	**GNG** (N=761)	**CSL** (N=803)
Intercept	182.34 (3.20)***	95.51 (1.55)***	116.30 (2.14)***	45.64 (0.87)***
PCL-M	-0.38 (0.11)***	-0.22 (0.05)***	-.015 (0.07)*	-0.11 (0.03)***
Male	9.42 (2.30)***	2.66 (1.11)	4.95 (1.53)**	1.12 (0.62)
Age: 20-24	5.00 (3.37)	-1.18 (1.63)	-3.31 (2.23)	-0.03 (0.92)
Age: 25-29	3.17 (3.50)	0.52 (1.69)	-2.92 (2.32)	0.49 (0.95)
Age: 30-34	-3.69 (3.66)	-2.63 (1.77)	-4.14 (2.41)	-2.26 (0.99)*
Age: 35-44	-7.00 (3.73)	-2.73 (1.80)	-7.40 (2.45)**	-4.17 (1.02)***
Age: 45-54	-15.32 (4.28)***	-11.11 (2.07)***	-21.37 (2.87)***	-9.68 (1.17)***
Age: 55-64	-23.27 (5.75)***	-16.23 (2.78)***	-25.40 (3.87)***	-12.60 (1.62)***
β	**CSR** (N=701)	**MTS** (N=759)	**SP** (N=798)	**SRT2** (N=793)
Intercept	46.62 (1.27)***	33.43 (0.87)***	33.55 (0.81)***	172.06 (3.46)***

PCL-M	-0.07 (0.04)	-0.04 (0.03)	-0.02 (0.03)	-0.61 (0.12)***
Male	1.65 (0.91)	1.96 (0.62)**	2.18 (0.58)***	9.65 (2.50)***
Age: 20-24	1.62 (1.32)	-0.80 (0.91)	-1.69 (0.58)*	1.92 (3.65)
Age: 25-29	1.01 (1.36)	0.69 (0.95)	-1.39 (0.87)	3.00 (3.80)
Age: 30-34	-0.29 (1.43)	0.20 (0.99)	-2.60 (0.93)**	2.87 (3.96)
Age: 35-44	-4.45 (1.45)**	-1.09 (1.00)	-4.23 (0.94)***	1.06 (4.03)
Age: 45-54	-10.54 (1.71)***	-4.95 (1.16)***	-7.57 (1.08)***	-3.68 (4.65)
Age: 55-64	-13.00 (2.39)***	-5.49 (1.60)***	-8.50 (1.45)***	-9.55 (6.20)

Note: $* = p < .05$; $** = p < .01$; $*** = p < .001$. The N for each subtest varies due to excluding administrations where less than 66 percent of trials were correctly completed.

Table 6: Comparisons of PCL-M-only models and fully specified models

Adjusted r^2

	PCL-M only	PCL-M + PSQI and PHQ-8	ANOVA	p
SRT1	0.08	0.08	$F(2, 793) = 1.56$	0.21
PRT	0.10	0.10	$F(2, 794) = 0.94$	0.39
GNG	0.12	0.12	$F(2, 750) = 0.15$	0.86
CSL	0.18	0.18	$F(2, 792) = 0.41$	0.67
CSR	0.13	0.12	$F(2, 690) = 0.36$	0.70
MTS	0.05	0.05	$F(2, 748) = 0.22$	0.80
SP	0.10	0.10	$F(2, 787) = 2.32$	0.10

| SRT2 | 0.05 | 0.05 | | F(2, 782) = 0.50 | 0.61 |

Figure 1: Age- and gender-adjusted slope effects of PCL-M score on throughput for subtests where the effect was significant.
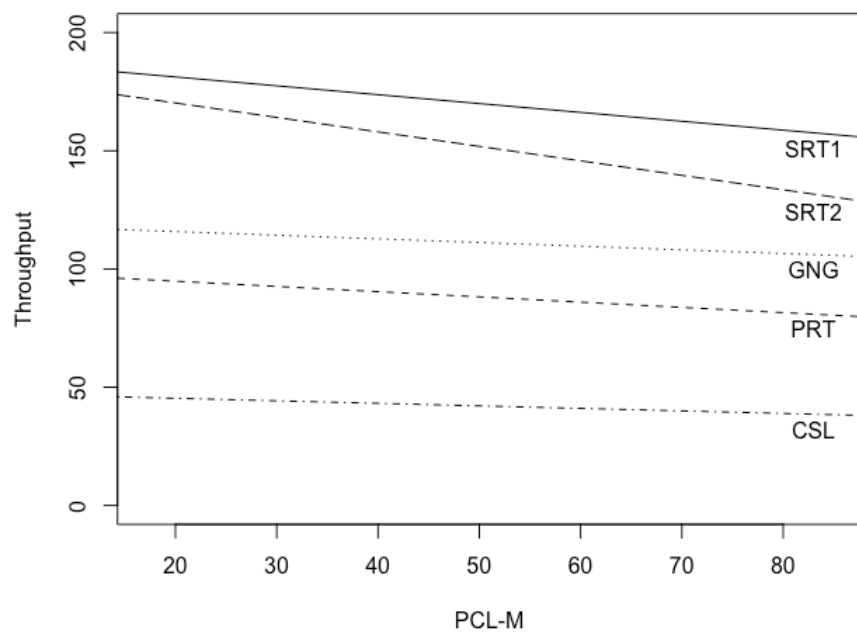
Note: PCL-M scores are presented on their original scale (17-85).
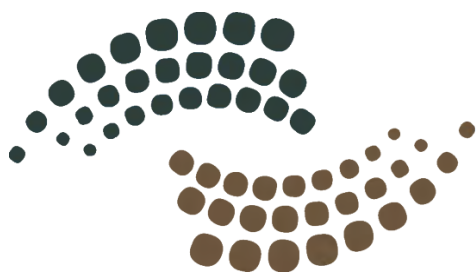
User Guide

version 1

for DANA v4.0.0-RIF

Army
RIF

Contract # W81XWH-12-C-0204

"Transitioning the Defense Automated Neurobehavioral Assessment (DANA) to Operational Use"

Contents

DANA Installation & Use

1

Install DANA on a Mobile Device

1. Enable **Developer options**:
    a. Go to *Settings* > About device and tap Build number 7 times.
        This should enable *Developer options*, creating it as a new *Settings* section.
        *On some devices, Developer options is located in *Settings* > *About device* > *Software info*.
2. Go to *Settings* > *Developer options* and turn on *USB debugging*.
3. Connect your device to your Windows or Mac PC via wired USB.
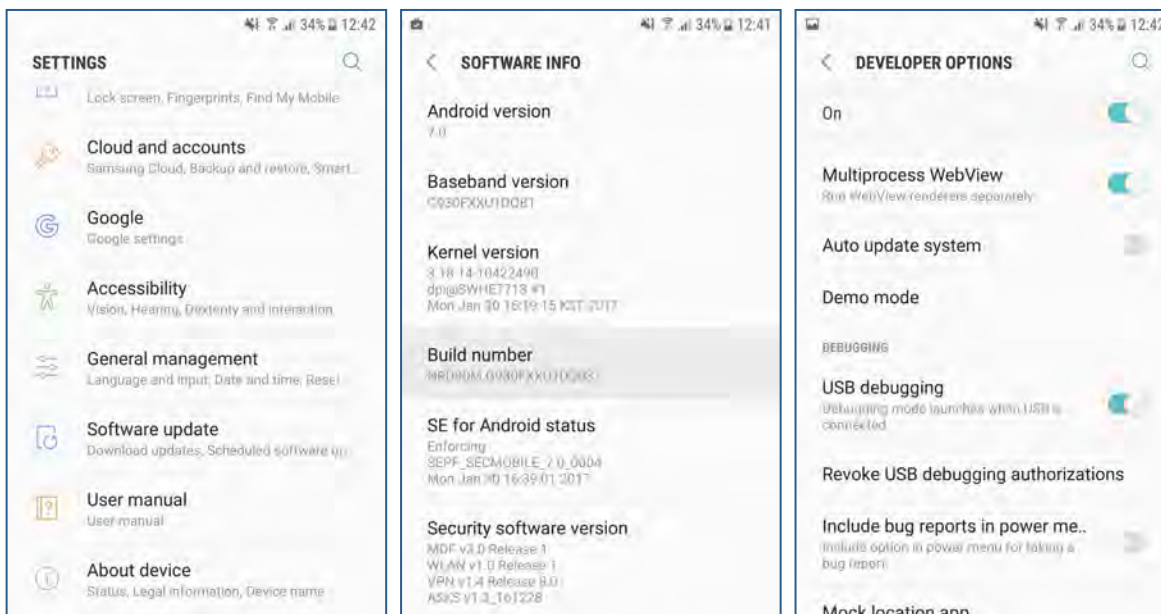
4. If prompted whether to *Allow USB debugging*, check the *Always allow from this computer* box and select *OK*.

5. Then, make sure MTP file transfer protocol is enabled:
   a. Depending on the device type, a notification may appear on-screen (see image below). If so, tap **ALLOW**.
   b. Otherwise, you may need to swipe down from the top of your screen to view your notifications, tap the USB notification, then **Transferring media files** (or the equivalent MTP connection choice).
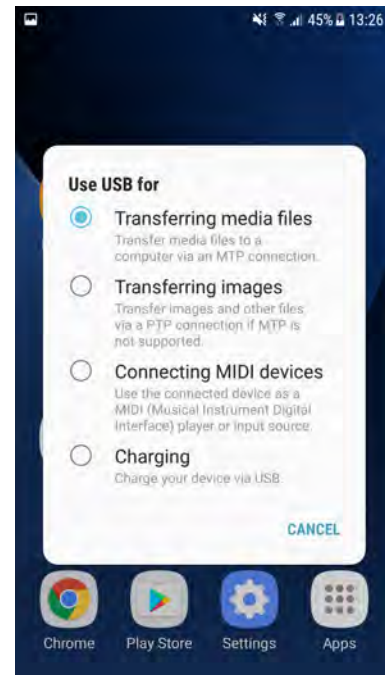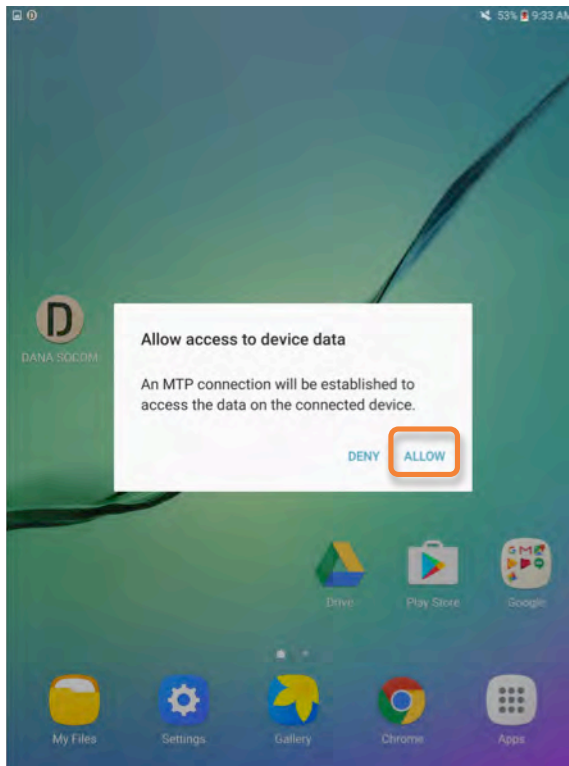
6. On the PC, double-click to open the mobile device in a file explorer window.
7. Copy the provided DANA install file (.apk format) to the mobile device.

Double-click to open
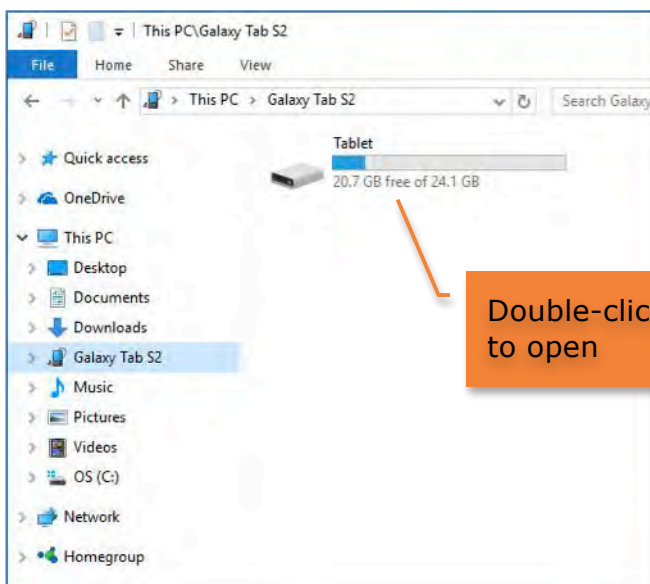
8. Select the apk file, then select **INSTALL**.
   The DANA app should then install on the mobile device.
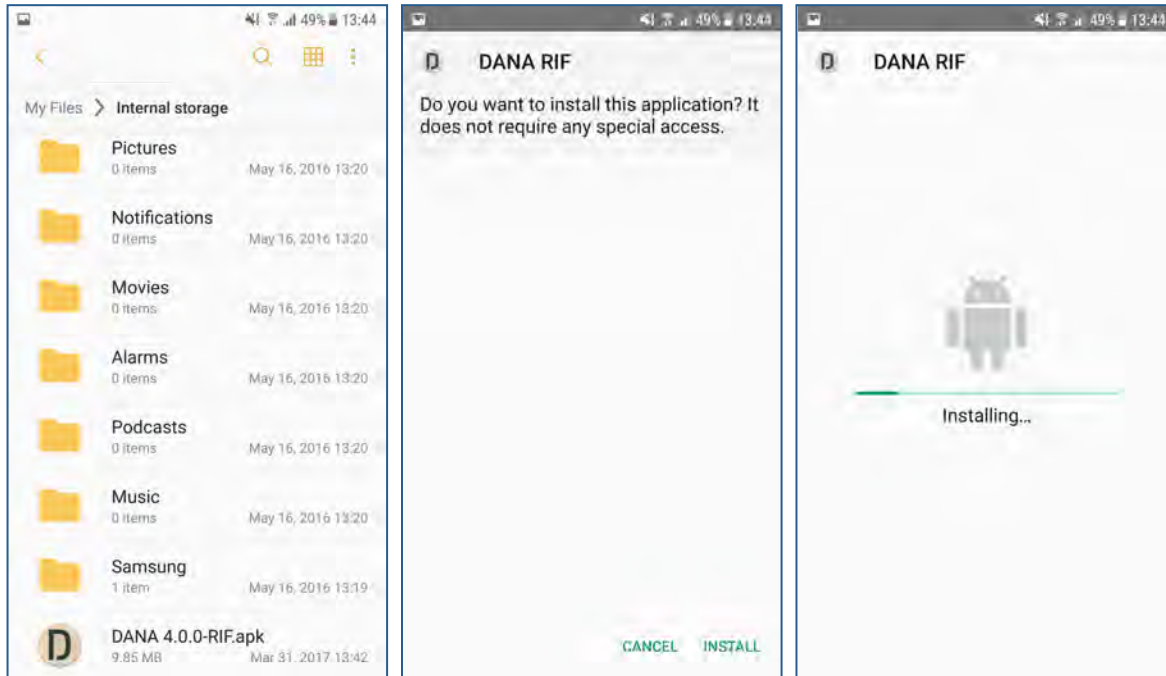
**If using a Mac PC*, make sure you have the Android File Transfer application installed:
https://www.android.com/filetransfer/

Once installed, Android File Transfer will automatically open when your mobile device is connected via USB and MTP is enabled, allowing you to continue with Step 7 above.

2 Open DANA

Launch the DANA RIF app from the Apps Tray.

If you are asked whether to "Allow DANA RIF to access photos, media, and files on your device," select **ALLOW**. This will allow the app to save exported data files to your device's internal storage.

If you selected **DENY** by accident, see **Appendix C** for instructions on how to change this setting.
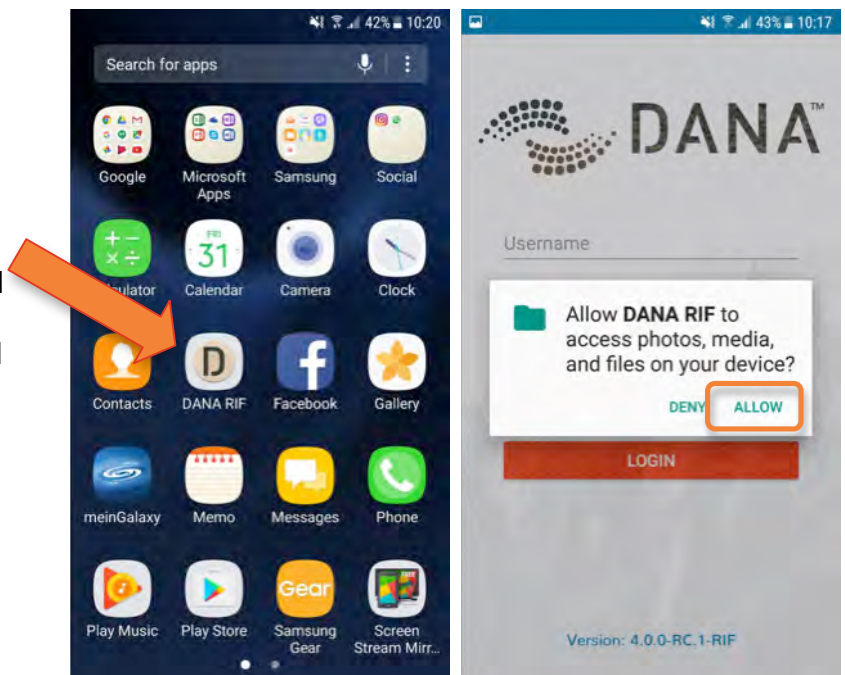
③ Log In

DANA RIF is designed to authenticate and upload data to the DANA cloud database (when the device is online). However, the cloud database endpoint has been removed from the app since its use requires the database to be hosted on a server connected to the Internet. Instructions for configuring this are included in a Readme file included along with the database and web portal repositories.

In lieu of online user authentication, one offline Administrator user has been hard-coded into the app. You can log in to the app as this user using these credentials:
- Username: admin
- Password: pass6677

DANA User Roles & Permissions

All DANA users fall into one of the following roles with the associated permissions:
- *Subject*: Can only take tests; cannot see any data
- *Examiner*: Can only see their subjects' results and all assessment options; can edit test batteries / individual tests
- *Clinician*: Can see all subjects' results and all assessment options; can edit test batteries / individual tests
- *Administrator*: Can see all subjects, results, and assessment options; can edit test batteries / individual tests; and can manage team members through the web portal

*Additional Examiner, Clinician, and Administrator users can be created via the Web Portal (assuming the cloud database and web portal are hosted on an internet-connected server). See **Section 9**.

*Note: Once logged in, if DANA is ever interrupted, you will be logged out for data security and privacy reasons. **If the interruption occurs during an assessment, data for that assessment will be discarded.***
Interruptions include:
- *Selecting the Home button*
- *Selecting the Multi-task button (and / or switching to another app)*
- *Putting the display to sleep (or allowing the display to timeout)*
- *Turning the device off*
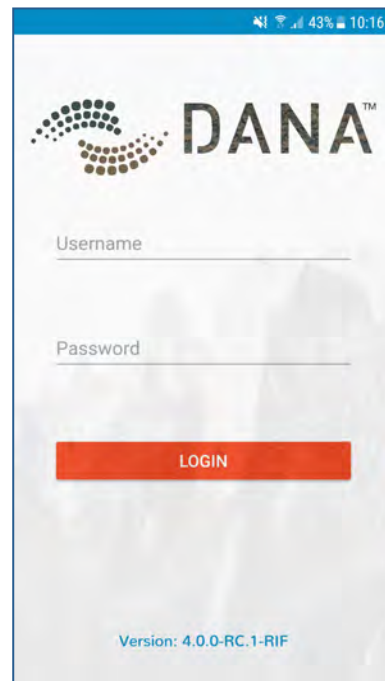
4 Add a New Subject

Navigate to the **Subjects** section of the app
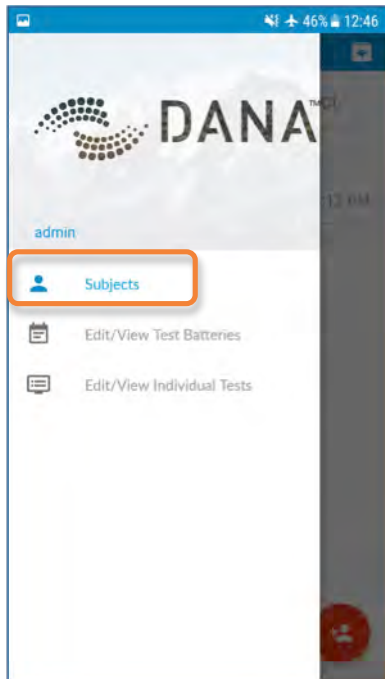
Select the **Add Subject** icon from the *Subjects* screen
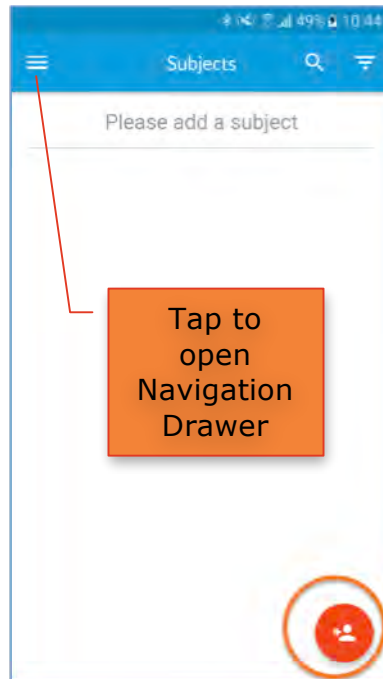
Enter a first name, last name, date of birth, then select **SAVE**

Tap to open Navigation Drawer
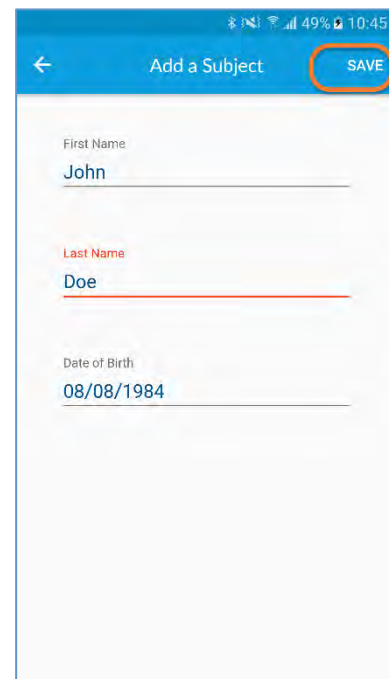
5 Select a Subject and an Assessment

Select a Subject
(or add a new one and select them)
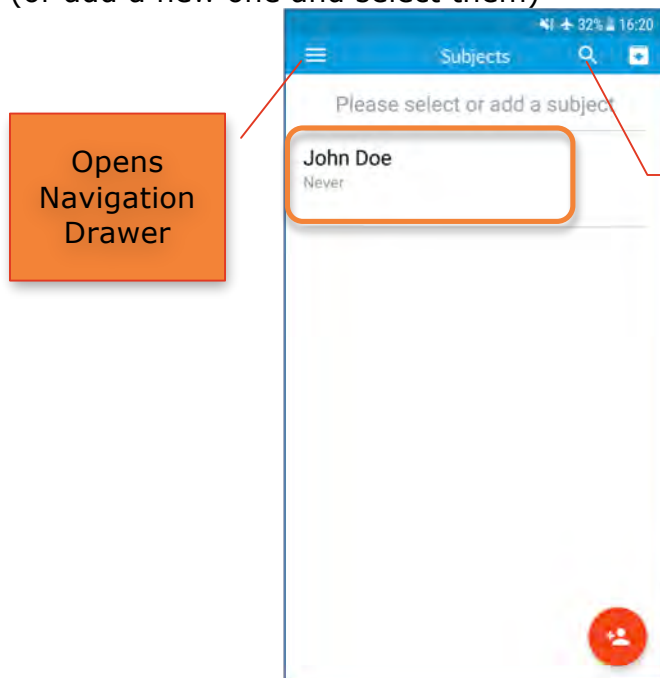
Select a test battery or an individual test to start an assessment

Opens Navigation Drawer

Search for a subject

Copyright 2017
AnthroTronix, Inc.
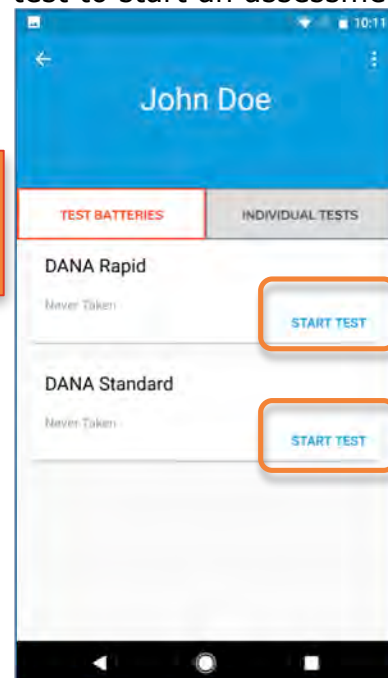
Default DANA Test Batteries

DANA Rapid

(~5 min)

1. Simple Reaction Time
2. Procedural Reaction Time
3. Go/No-Go

DANA Standard

(~20 min)

1. Simple Reaction Time
2. Code Substitution (Learning)
3. Procedural Reaction Time
4. Spatial Processing
5. Go/No-Go
6. Match to Sample
7. Memory Search (Sternberg)
8. Simple Reaction Time
9. Patient Health Questionnaire 8
10. Insomnia Severity Index

Default DANA Individual Tests

1. Simple Reaction Time
2. Procedural Reaction Time
3. Go/No-Go
4. Code Substitution (Learning & Recall)
5. Spatial Processing
6. Match to Sample
7. Memory Search (Sternberg)
8. Patient Health Questionnaire 8
9. Insomnia Severity Index
10. PTSD Check List – Civilian
11. Primary Care PTSD screen
12. Pittsburgh Sleep Quality Index
13. Stanford Sleepiness Scale

View Results on the Mobile Device

Log into DANA
(if logged out)
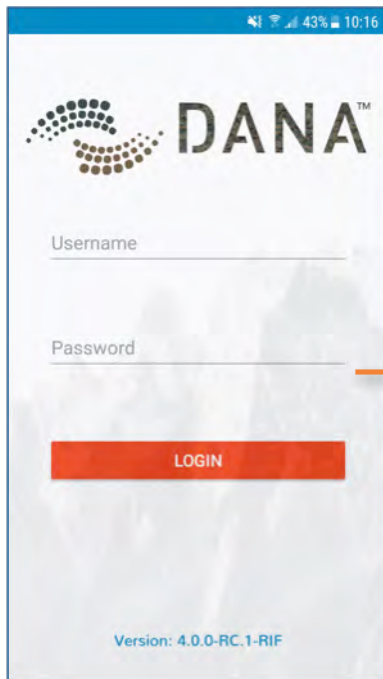
After selecting a subject, select **VIEW RESULTS** for the desired assessment type
Then select **View Result** for the specific date and time
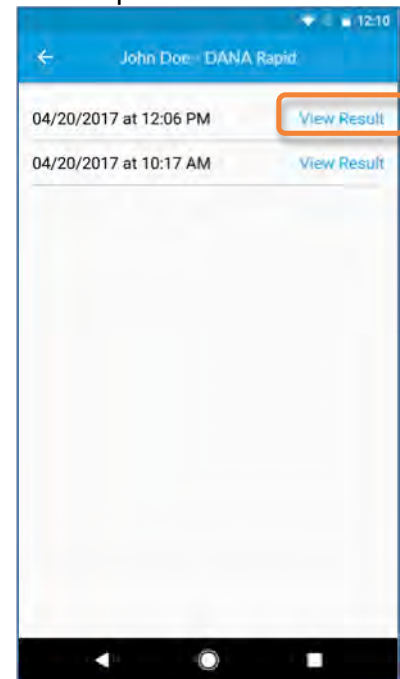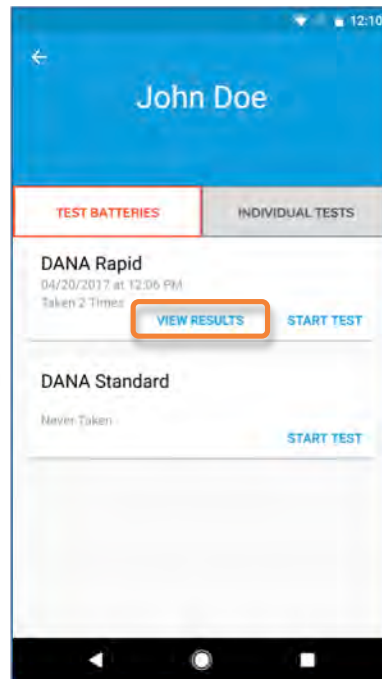
Choose how to view your results using the tabs at the top:

Summary	Graph	Raw Data

On small screens: tap here to toggle to Mean Reaction Time & % Correct

Use the drop down menu to choose which individual test's results to view. Use the bottom tabs to choose graph.

7 Export Results Data from the DANA App

Select the *Global Export* button to export all
test results on that device in CSV format
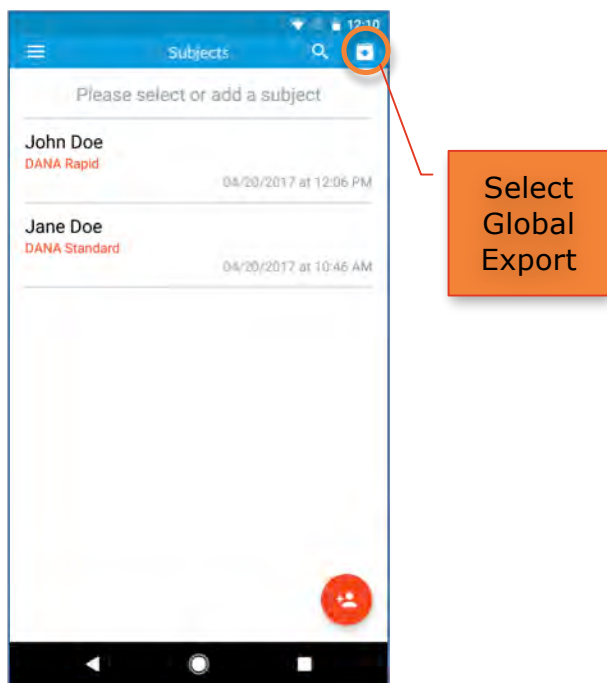
All data on that mobile device will then be saved to the *DANA > Exports* directory in the device's
storage.

The CSV files can then be transferred to a PC via a wired USB connection (see **Section 10**).
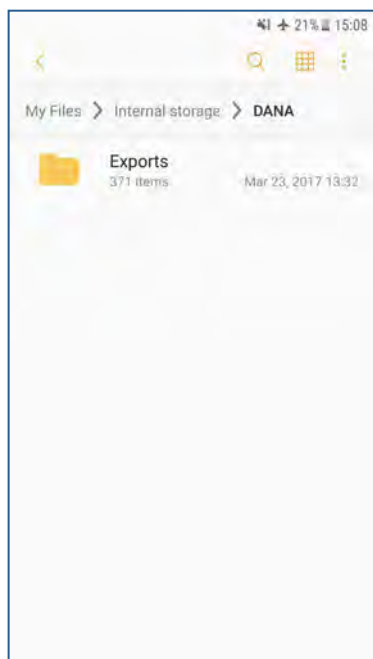
Edit Individual Tests and Test Batteries

Choose **Edit/View Test Batteries** or **Edit/View Individual Tests** from the navigation drawer
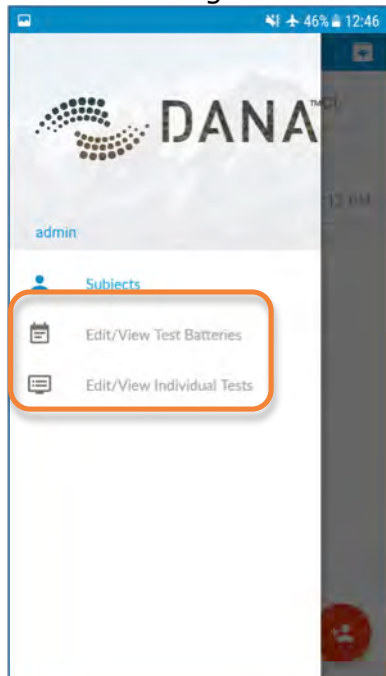
Test Batteries:
Select a battery to edit or select the plus sign to create a new one
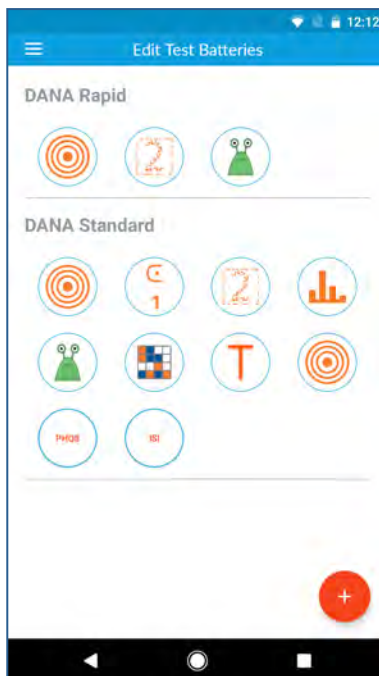
Test Batteries:
Select tests to add, then select **Add Test** to add them to the battery

Test Batteries:

You can create a new or modify an existing battery
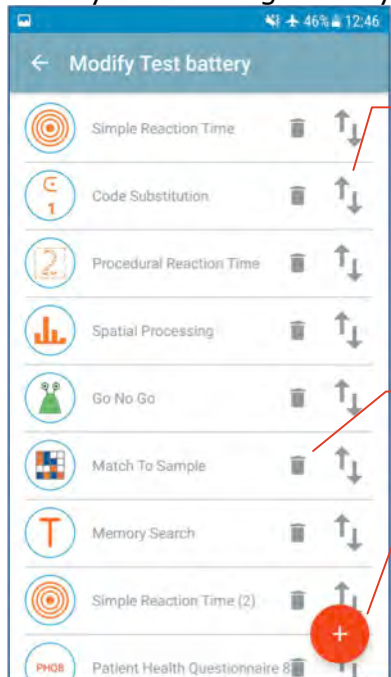
Reorder Test (hold & move)

Delete Test

Add Test

Individual Tests:
You can select an an individual test and modify its parameters
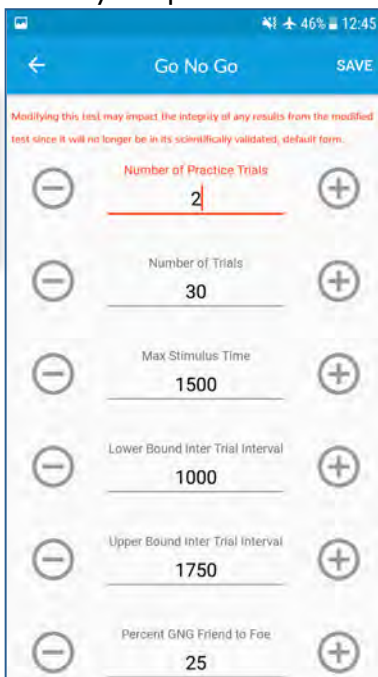
Save each new test or battery with a unique name
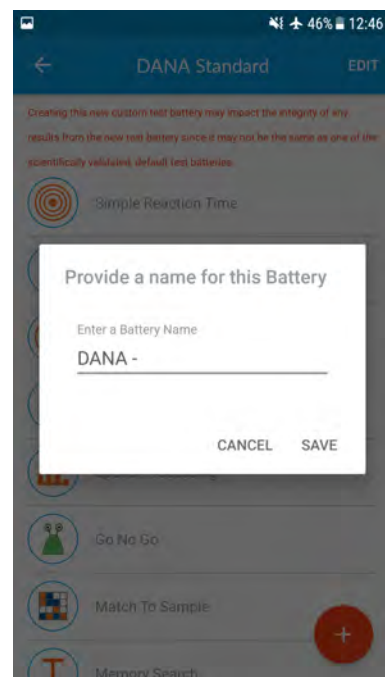
Code Substitution: Learning vs. Recall section

The Code Substitution test has two potential sections: Learning and Recall. The Learning section teaches a code set to the Subject; the Recall section then tests the Subject's ability to recall that same code set.

By default, Code Substitution is the Learning section. To add a Recall section to a test battery, add a second (or more) instance(s) of Code Substitution to the test battery (as shown below).

When one or more Recall sections follows the Learning section, the Learning section increases (by default) from 36 to 72 regular trials.

Copyright 2017
AnthroTronix, Inc.
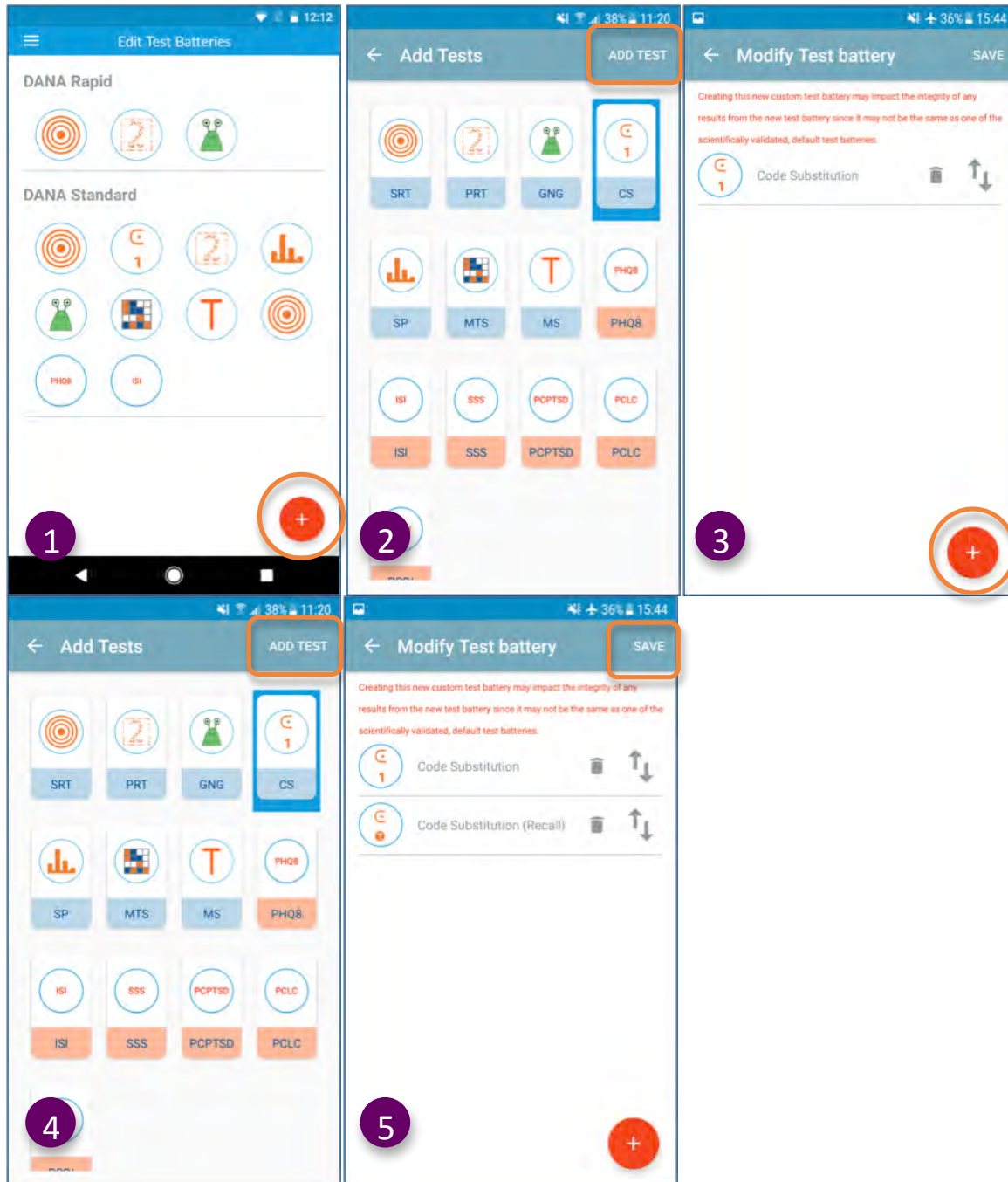
DANA Web Portal

If both the web portal and cloud database code is hosted on an internet-connected server and endpoints are specified in the app, the following instructions would apply for web portal use:

Visit *[web portal URL – to be specified]* and log in

After logging in, you can access*…

*See **Section 3** for permissions by user role

1. Subjects section
- View subjects list
- Select a subject to view assessments

2. Results section
- View a subject's assessments
- Select an assessment to view results
- Results can also be downloaded in both PDF and CSV format

3. Manage Team section
- Create new users
- Manage members of the administrative team

1. Make sure that USB debugging is turned on. (See instructions in **Section 1**.)
2. Connect your device to your Windows or Mac PC via wired USB.
2. Then, make sure MTP file transfer protocol is enabled:
   - Depending on the device type, a notification may appear on-screen (see image below). If so, tap **ALLOW**.
   - Otherwise, you may need to swipe down from the top of your screen to view your notifications, tap the USB notification, then **Transferring media files** (or the equivalent MTP connection choice).
3. On the PC, double-click to open the mobile device in a file explorer window and navigate to the directory in *[Device] > DANA > Exports*. All exported files should be located in this directory.
4. Copy exported files to the PC.
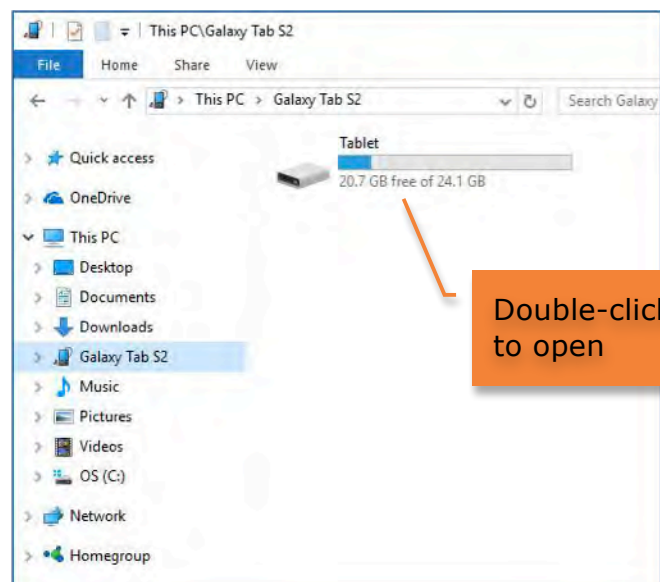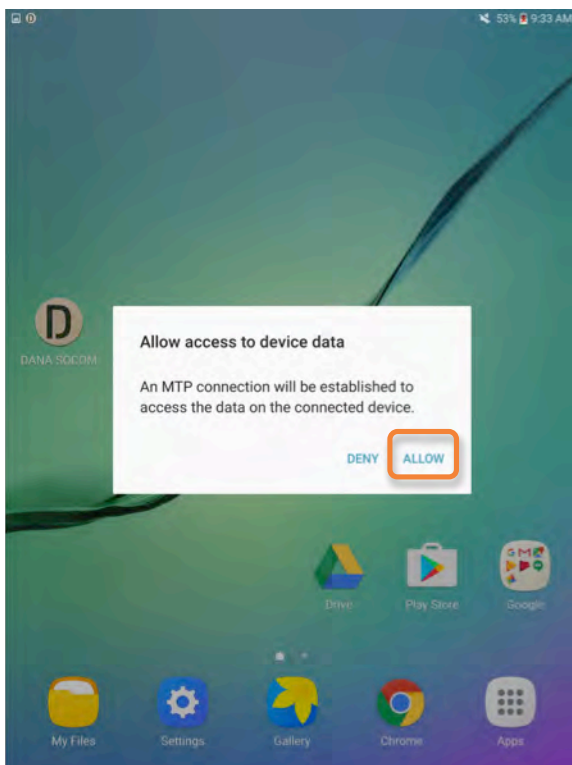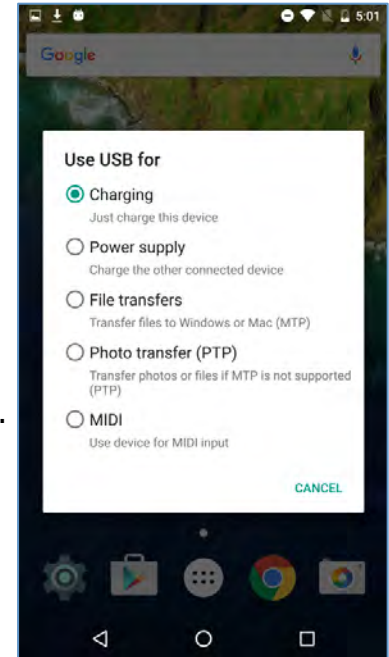
Double-click to open

***If using a Mac PC**, make sure you have the Android File Transfer application installed: https://www.android.com/filetransfer/

Once installed, Android File Transfer will automatically open when your mobile device is connected via USB and MTP is enabled; navigate to the DANA > Exports directory and then files to the PC.
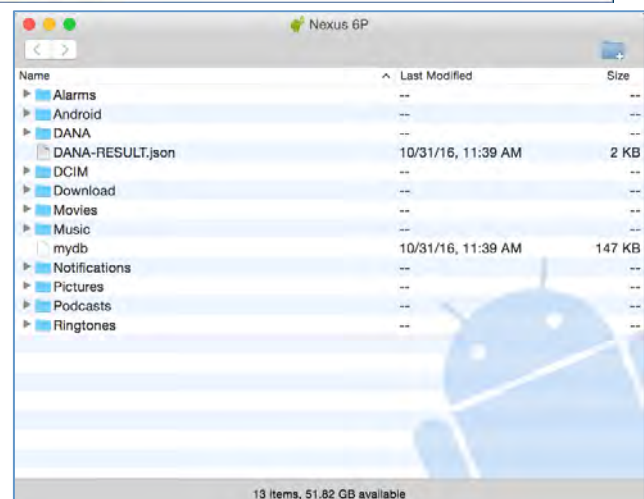
Interpreting the Results Screen

In both the Android DANA application and the DANA web portal, the number of correctly completed trials on each test is summarized via a three-way categorization scheme (✓, -, or x):

- ✓ **Acceptable number of trials completed**: the subject has completed an acceptable number of trials, i.e., a number not affected by either of the constraints described below.
- - **Unacceptable number of trials completed**: the number of trials completed by the subject is at or below the 5th percentile of normal performance. See the appendix on page 18 for a description of the normative reference data used.
- X **Test incomplete**: the subject correctly completed less than 66% of trials, which is at or near chance performance depending on the subtest.
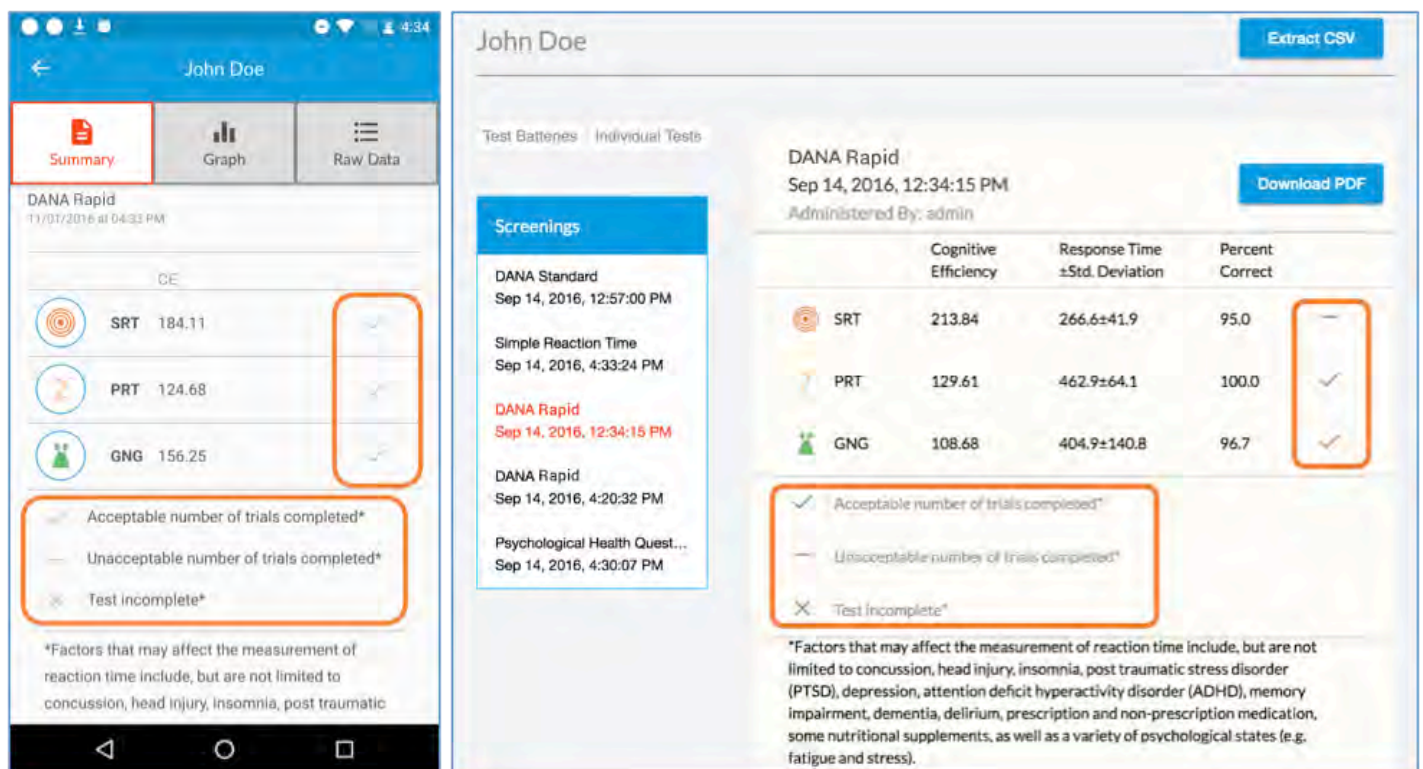
The summary screens also suggest factors that may affect performance, such as head injury, memory impairment, dementia, etc. However, this list is not exhaustive; some factors such as age (e.g., very young or very old) are also likely to have performance consequences. DANA administrators should rely on their judgment and experience when considering possible causes of a less than ideal number of trials correctly completed.
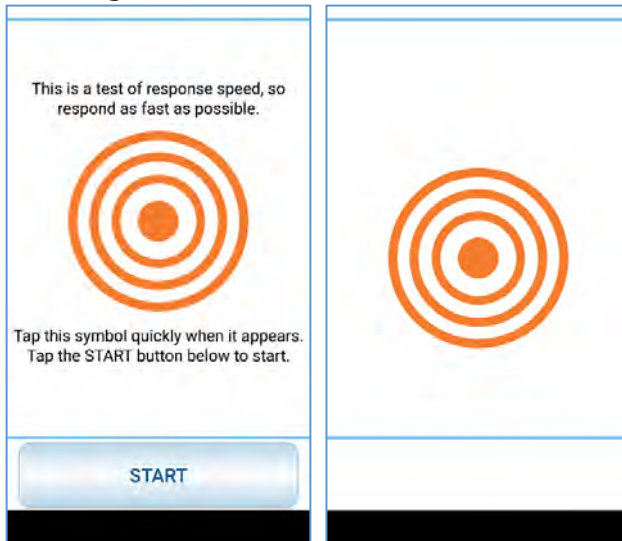
DANA Test & Test Battery Descriptions

DANA includes both cognitive tests and psychological surveys. The default configuration for each is represented below.

Simple Reaction Time

The subject taps on the location of the orange target symbol as quickly as possible each time it appears.
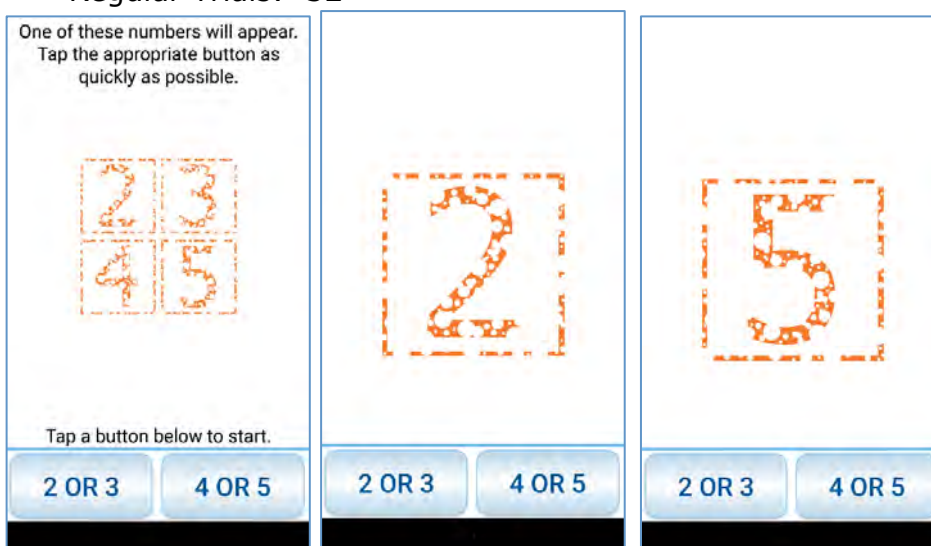
Practice Trials: 5
Regular Trials: 40

This is a test of response speed, so respond as fast as possible.

Tap this symbol quickly when it appears. Tap the START button below to start.

START

Procedural Reaction Time

The screen displays one of four numbers (2, 3, 4 or 5) for 2 seconds. The subject taps the left button (**2 or 3**) or right button (**4 or 5**) at the bottom of the screen as quickly as possible to indicate which number was displayed.

Practice Trials: 10
Regular Trials: 32

One of these numbers will appear. Tap the appropriate button as quickly as possible.

Tap a button below to start.

| 2 OR 3 | 4 OR 5 | 2 OR 3 | 4 OR 5 | 2 OR 3 | 4 OR 5 |

Go/No-Go

A house is presented on the screen with several windows. Either a "friend" (green) or "foe" (gray) appears in a window. The subject must tap the **BLAST** button only when a "foe" appears.
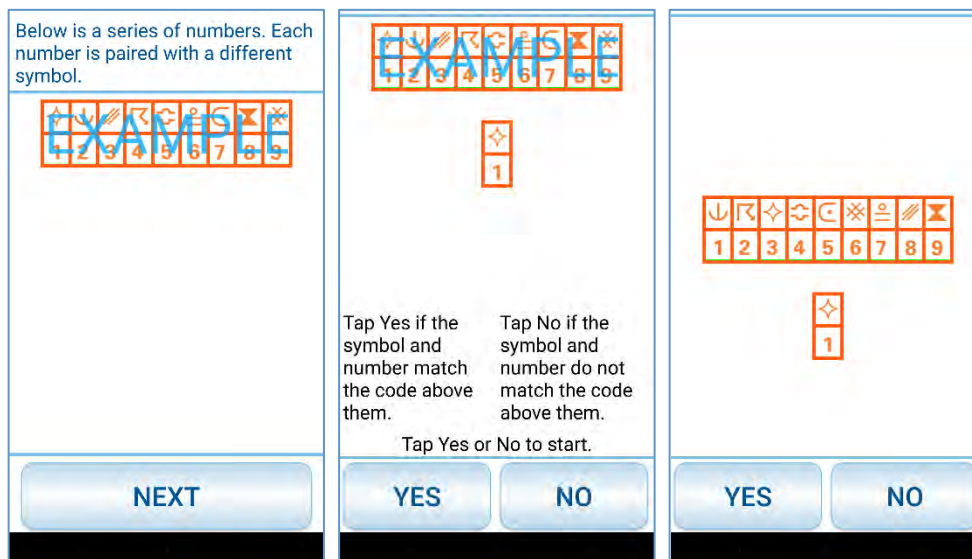
Practice Trials: 5
Regular Trials: 30

An alien will appear in a window of this building.	**NEXT**
This alien is a foe. Tap the BLAST button if the alien is a foe. / This alien is a friend. Do nothing if the alien is a friend. / Tap BLAST to start.	**BLAST**
	BLAST
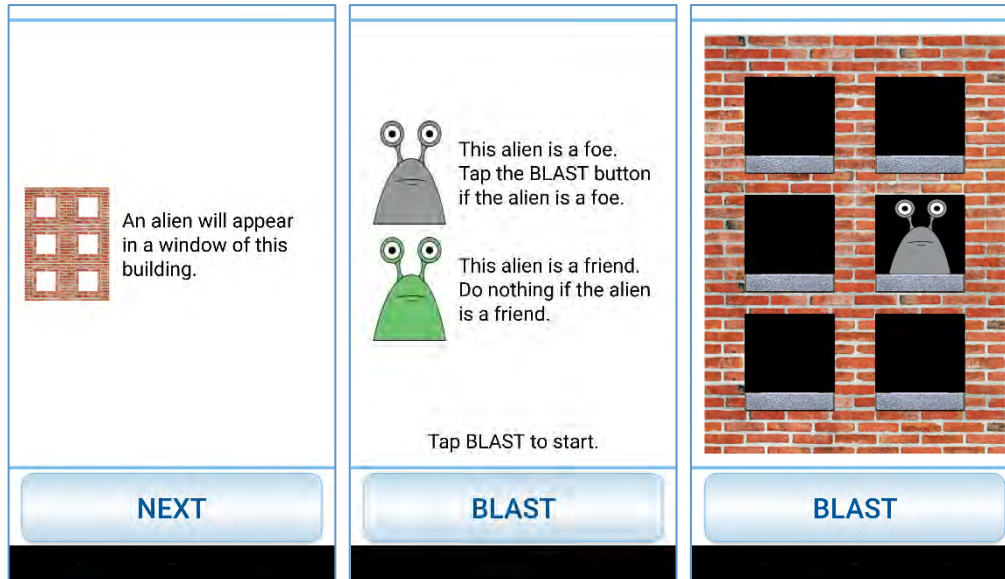
Code Substitution (Learning Section)

Subjects refer to a code set of 9 symbol-digit pairs that is shown on the screen. Single symbol-digit pairs are presented in succession below the code, and the subject indicates whether or not the single pair matches the code by tapping **YES** or **NO**.

Practice Trials: 4
Regular Trials (if no Recall section after): 36
Regular Trials (if Recall section after): 72

Below is a series of numbers. Each number is paired with a different symbol.

Tap Yes if the symbol and number match the code above them.

Tap No if the symbol and number do not match the code above them.

Tap Yes or No to start.

Code Substitution (Recall Section)

After a delay of several intervening tests, single symbol-digit pairs are presented again; this time without the code above. The subject indicates whether or not the pairing was included in the code that was presented in the earlier Code Substitution (Learning) section by tapping **YES** or **NO**.

Practice Trials: 0
Regular Trials: 36

Now you will be asked to remember the symbol / number pairs you learned earlier.
During this section a symbol and number will appear, but this time the code will not be shown.

NEXT

Tap Yes if the symbol and number match the code you saw earlier.

Tap No if the symbol and number do not match the code you saw earlier.

Tap Yes or No to start.

YES   NO

YES   NO

Spatial Processing

Pairs of four-bar histograms are displayed on the screen simultaneously, one rotated 90 degrees (either clockwise or counterclockwise). The subject is required to determine whether they would identical if the rotation was not applied.
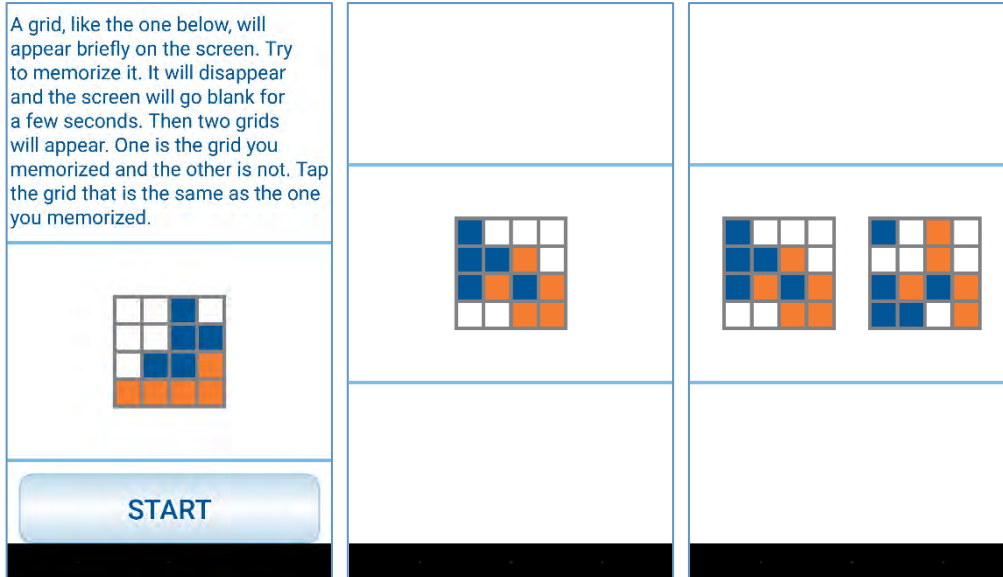
Practice Trials: 10
Regular Trials: 20

Two bar graphs will appear on the screen.
Tap the "Same" button if when rotated upright the graphs would be the same.
Tap the "Different" button if when rotated upright the graphs would not be the same.

Tap Same or Different to start.

SAME   DIFFERENT

SAME   DIFFERENT

SAME   DIFFERENT

Match to Sample

A single 4 x 4 checkerboard pattern is presented on the screen for brief study period. It then disappears for 5 seconds, after which two patterns are presented side-by-side. The subject indicates which of these two patterns was displayed during the study period.
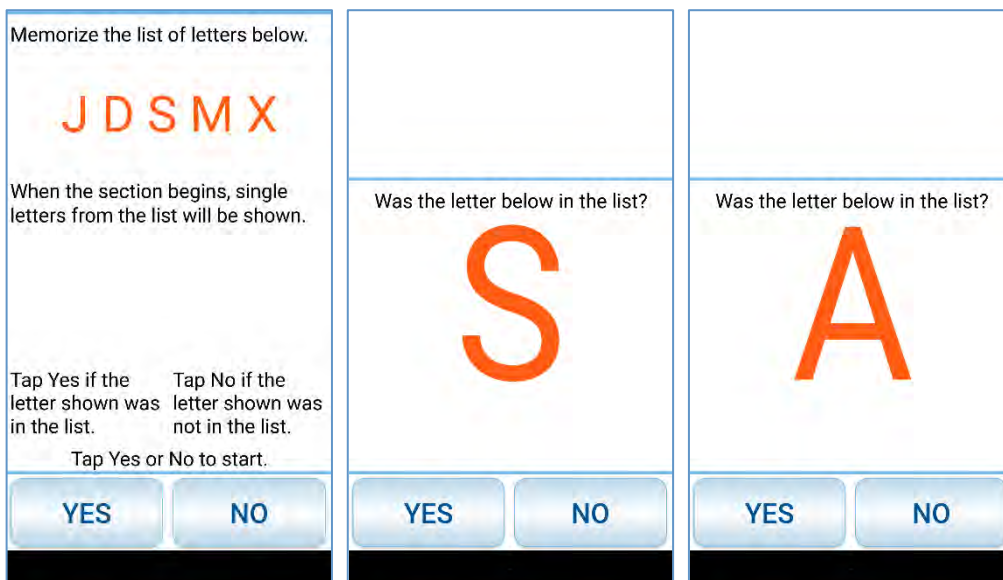
Practice Trials: 3
Regular Trials: 20

A grid, like the one below, will appear briefly on the screen. Try to memorize it. It will disappear and the screen will go blank for a few seconds. Then two grids will appear. One is the grid you memorized and the other is not. Tap the grid that is the same as the one you memorized.

START

Memory Search (Sternberg)

Before the test begins, the subject is required to memorize a list of five letters. Then, each trial presents a single letter, and the subject must indicate whether that letter is contained in the memorized list.

Practice Trials: 0
Regular Trials: 30

Memorize the list of letters below.

J D S M X

When the section begins, single letters from the list will be shown.

Tap Yes if the letter shown was in the list. Tap No if the letter shown was not in the list.

Tap Yes or No to start.

YES   NO

Was the letter below in the list?

S

YES   NO

Was the letter below in the list?

A

YES   NO

Patient Health Questionnaire 8 (PHQ-8)

Purpose: Measures depression
Questions: 8-9

Over the last 2 weeks, how often have you been bothered by any of the following problems?	Over the last 2 weeks, how often have you been bothered by any of the following problems?	Over the last 2 weeks, how often have you been bothered by any of the following problems?
Trouble falling or staying asleep, or sleeping too much.	Feeling down, depressed, or hopeless.	Little interest or pleasure in doing things.
Not at all	Not at all	Not at all
Several days	Several days	Several days
More than half the days	More than half the days	More than half the days
Nearly every day	Nearly every day	Nearly every day
NEXT	NEXT	NEXT

Primary Care – PTSD Screen (PC-PTSD)

Purpose: Initial screen for PTSD to be used in primary care
Questions: 4

	In your life, have you ever had any experience that was so frightening, horrible, or upsetting that, in the past month, you …	In your life, have you ever had any experience that was so frightening, horrible, or upsetting that, in the past month, you …
	Were constantly on guard, watchful, or easily startled?	Have had nightmares about it or thought about it when you did not want to?
The following questions relate to previous difficult experiences, and their effect on your life in the past month.	Yes	Yes
	No	No
NEXT	NEXT	NEXT

Insomnia Severity Index (ISI)

Purpose: Measures severity of insomnia
Questions: 7

The following section will ask you a series of questions about any sleep problems. Please read each one carefully, and select the appropriate response.	Please rate the current (i.e., in the past week) severity of your insomnia problem(s). Problem waking up too early?	Please rate the current (i.e., in the past week) severity of your insomnia problem(s). Difficulty falling asleep?
	None	None
	Mild	Mild
	Moderate	Moderate
	Severe	Severe
	Very severe	Very severe
NEXT	NEXT	NEXT

Pittsburgh Sleep Quality Index (PSQI)

Purpose: Measures quality of sleep
Questions: 19-24

The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month.	During the past month, what time have you usually gone to bed?	During the past month, how often have you had trouble sleeping because you … Wake up in the middle of the night or early morning
	11:00 AM PM	Not during the past month
	11 12 1 10 2 9 3 8 4 7 6 5	Less than once a week
		Once or twice a week
		Three or more times a week
NEXT	**NEXT**	NEXT

PTSD Check List for Civilians (PCL-C)

Purpose: Measures PTSD-related symptoms
Questions: 17

The following will present a list of problems and complaints that people sometimes have in response to stressful life experiences. Please read each one carefully, select an option to indicate how much you have been bothered by that problem in the last month.

NEXT

In the past month, how much have you been bothered by

Repeated, disturbing memories, thoughts, or images of a stressful experience from the past?

Not at all

A little bit

Moderately

Quite a bit

Extremely

NEXT

In the past month, how much have you been bothered by

Repeated, disturbing dreams of a stressful experience from the past?

Not at all

A little bit

Moderately

Quite a bit

Extremely

NEXT

Stanford Sleepiness Scale (SSS)

Purpose: Measures sleepiness
Questions: 1

How sleepy do you feel today on a scale of 1 – 7?

1. Feeling active, vital, alert, or wide awake

2. Functioning at high levels, but not at peak; able to concentrate

3. Awake, but relaxed; responsive but not fully alert

4. Somewhat foggy, let down

5. Foggy; losing interest in remaining awake; slowed down

6. Sleepy, woozy, fighting sleep;

More Answers
Drag the screen up or down to view

More Answers (Down)

NEXT

How sleepy do you feel today on a scale of 1 – 7?

More Answers (UP)

2. Functioning at high levels, but not at peak; able to concentrate

3. Awake, but relaxed; responsive but not fully alert

4. Somewhat foggy, let down

5. Foggy; losing interest in remaining awake; slowed down

6. Sleepy, woozy, fighting sleep; prefer to lie down

7. No longer fighting sleep, sleep onset soon; having dream-like thoughts

NEXT

Appendix A: Description of Normative Reference Data Used for the Summary Screen

The "✓," "-," and "X" that categorize the number of correctly completed trials on the summary screen are based in part on DANA data from a sample of 552 healthy U.S. military service members aged 18 – 64. The "-" designation is applied to subtest administrations where the proportion of correctly completed trials is below the 5th percentile as determined by this normative dataset.

The 5th percentile was calculated after excluding administrations where less than 66% of trials were correctly completed (i.e., administrations with the "X" designation). The table below presents percentile distributions for percent correct for each subtest.

Subtest	5%	25%	50%	75%	95%
SRT1	95.0	100.0	100.0	100.0	100.0
PRT	93.8	96.9	100.0	100.0	100.0
GNG	93.3	100.0	100.0	100.0	100.0
CSL	87.5	94.4	97.2	98.6	100.0
CSR	69.4	80.6	88.9	94.4	100.0
MTS	73.3	80.0	86.7	93.3	96.7
SP	80.0	90.0	95.0	95.0	100.0
MS*	80.0	93.3	96.7	100.0	100.0
SRT2	92.5	97.5	100.0	100.0	100.0
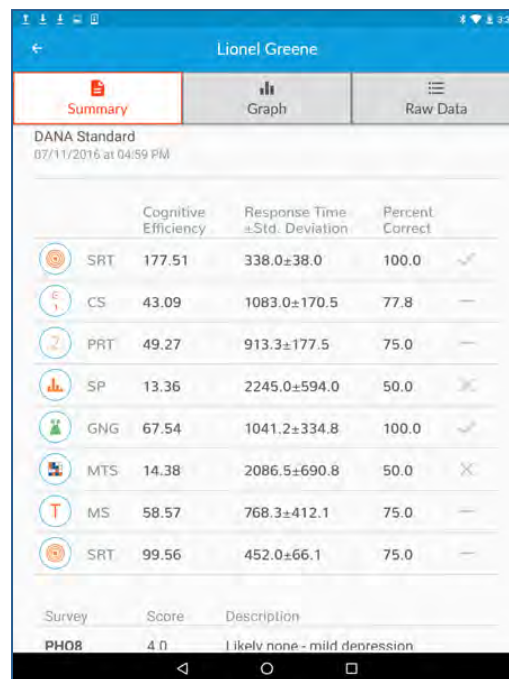
* Data for the Memory Search (MS) subtest were not collected from the sample described above. Normative values for this subtest were collected from a demographically similar sample of 124 healthy adults.

Appendix B: Description of DANA Summary Statistics

Three primary summary statistics are calculated by DANA and displayed in report screens: (1) Cognitive Efficiency, (2) Mean Correct Response Time (+/- a standard deviation), and (3) Percent Correct.

Cognitive Efficiency:
Cognitive Efficiency (CE), a combined measure of both speed and accuracy, is the amount of correct responses per minute. Units: correct responses / minute

$$CE = \left(\frac{percentCorrect}{meanCorrect\,\mathrm{Re}\,actionTime}\right) \times 60,000$$

Mean Correct Response / Reaction Time:
Represented on report screens as simply Response Time or Reaction Time, this value is the average (mean) response time for all correct trials for a given test administration.
For the Go/No-Go test only, response times for correct "No-Go" trials are excluded from this calculation. Units: milliseconds (ms)

$$meanCorrectRT = average(correctRT_1, correctRT_2, \ldots, correctRT_n)$$

Percent Correct:
This value is simply the percentage of correct trials for a given test administration. Units: %

$$PC = \frac{numberOfCorrectTrials}{totalNumberOfTrials} \times 100\%$$

Copyright 2017
AnthroTronix, Inc.
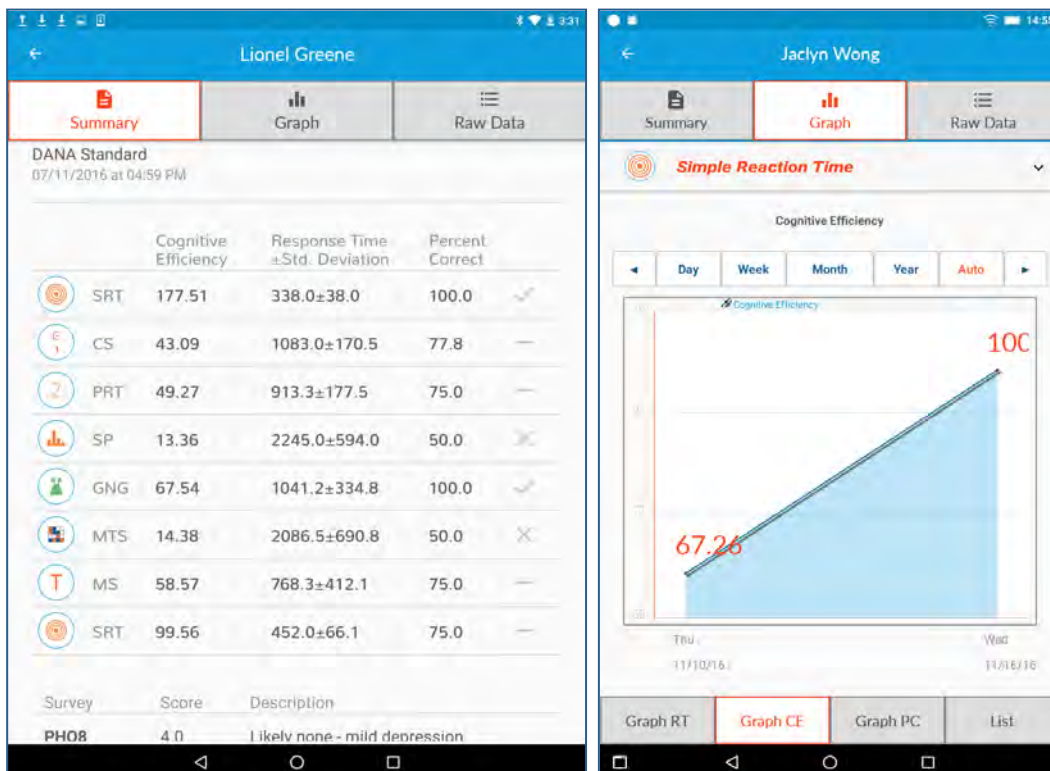
Appendix C: Adjusting App Permissions

The first time the DANA RIF app is opened, a permissions prompt should appear on the login screen (see screenshot) asking if the app may have access to "photos, media, and files on your device." If **DENY** was selected in that prompt, the app will not be able to save exported CSV data to the device's internal storage. But you can change this permissions setting to grant access to device files and allow exported CSV data files to be saved:

1. Go to the Settings app on your device

2. Select **Applications**, then **Application manager** (if needed), then the *DANA RIF* app
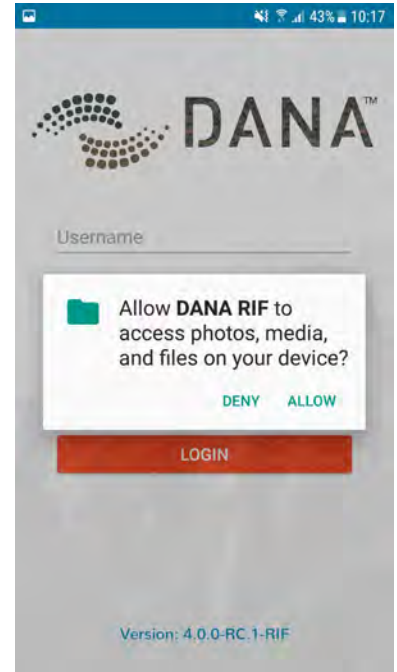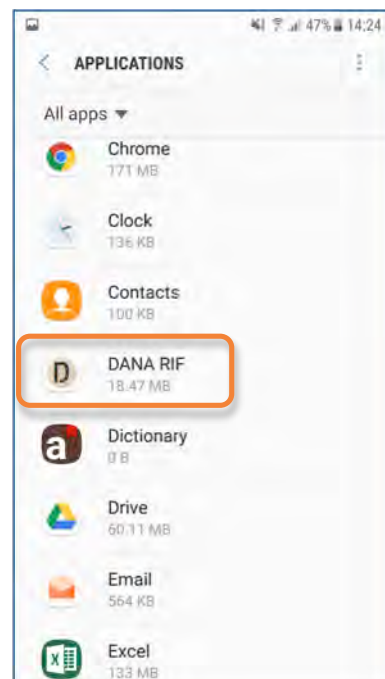
3. Select Permissions, then toggle the **Storage** permission to *ON*.

The DANA app should now be able to save exported data files to the device's internal storage.

DANA™ Technology Transition Package

I.   Description of DANA

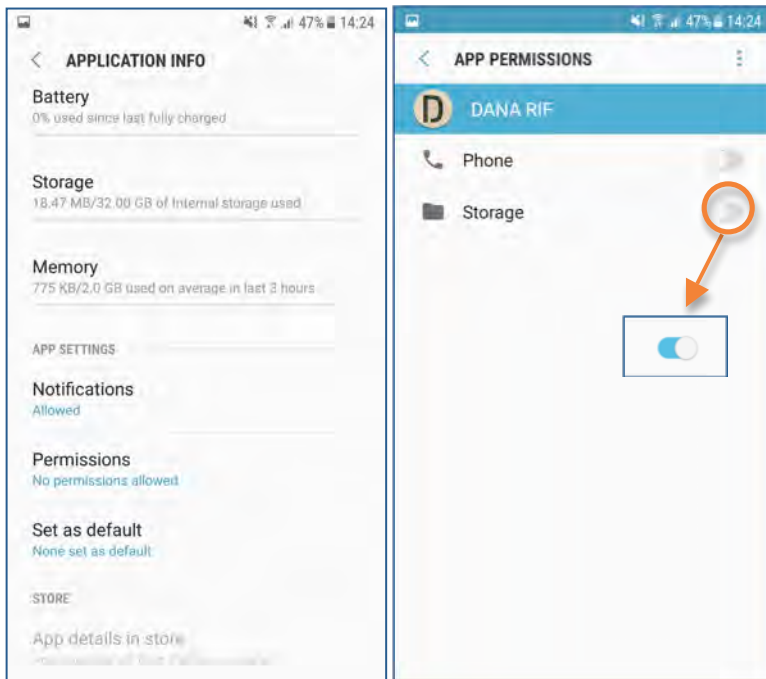Developed by AnthroTronix, Inc. (ATinc), DANA™, the Defense Automated Neurobehavioral Assessment, is an FDA-cleared, Neurocognitive Assessment Test (NCAT) running as a mobile application on Android-based devices. DANA has been demonstrated as valid and reliable in all settings, and has shown sensitivity and distinguished impaired vs. non-impaired in a variety of populations, including extremely depressed, acute concussion, hypoxia, and mild cognitive impairment due to PTSD. DANA has been successfully administered to members of the Army, Navy, Air Force, and Marines in five extreme environments: arctic, altitude, jungle, desert, and shipboard in high sea states.

Included in DANA are assessments of speed and accuracy on a number of standardized neurocognitive tests, along with standardized psychological assessments (e.g., for depression, PTSD, disturbed sleep, etc.). Test results are available immediately and are displayed on intuitive reporting screens that make it easy to track an individual's performance over time; they can also be easily exported in CSV and PDF file formats.

These individual tests are combined into two standard test batteries: DANA Brain Vital, which can be administered in under five minutes, and DANA Standard, which takes approximately 20 minutes to administer, as shown in Table 1 below. In addition, DANA Modular enables clinicians to customize test batteries in real-time, and which can include any combination of individual cognitive and psychological tests, and parameters within each of those tests, such as the number of trials, can be modified as well.

Because DANA assesses the speed of response, as well as response accuracy, ATinc has established specifications for the maximum variability in latency between when a user touches a screen in response to a stimulus, and when the system records the response. ATinc has conducted extensive testing to determine if specific Android-based devices meet this specification.

Figure 1 below shows the screen shots from Simple Reaction Time (SRT), Procedural Reaction Time (PRT), and Go/No-Go Decision Making Tests (GNG), and Figure 2 shows sample report screens.

Table 1: DANA Test Battery Configurations

DANA Brain Vital (5 minutes)	DANA Standard (20 minutes)	DANA Modular (Varies)
Simple Reaction Time	Simple Reaction Time	Can be configured to include any combination of individual cognitive and psychological tests.
Procedural Reaction Time	Code Substitution (Learning)	
Go/No-Go	Procedural Reaction Time	
	Spatial Reasoning	Individual cognitive tests can be modified to change parameters, including the number of trials.
	Go/No-Go	Additional tests available include: • Combat MACE interview • Combat Exposure Scale (CES) • PTSD Check List – Military Version (PCL-M) • Deployment Stress Inventory (DSI) • Code Substitution (Recall) • Pittsburgh Sleep Quality Index (PSQI) • Primary Care PTSD screen (PC-PTSD) • Stanford Sleepiness Scale
	Matching to Sample	
	Memory Search (Sternberg)	
	Simple Reaction Time	
	Patient Health Questionnaire (PHQ-8)	
	Insomnia Severity Index (ISI)	

Figure 1: Screen Shots from Simple Reaction Time, Procedural Reaction Time, and Go/No-Go Decision Making Tests
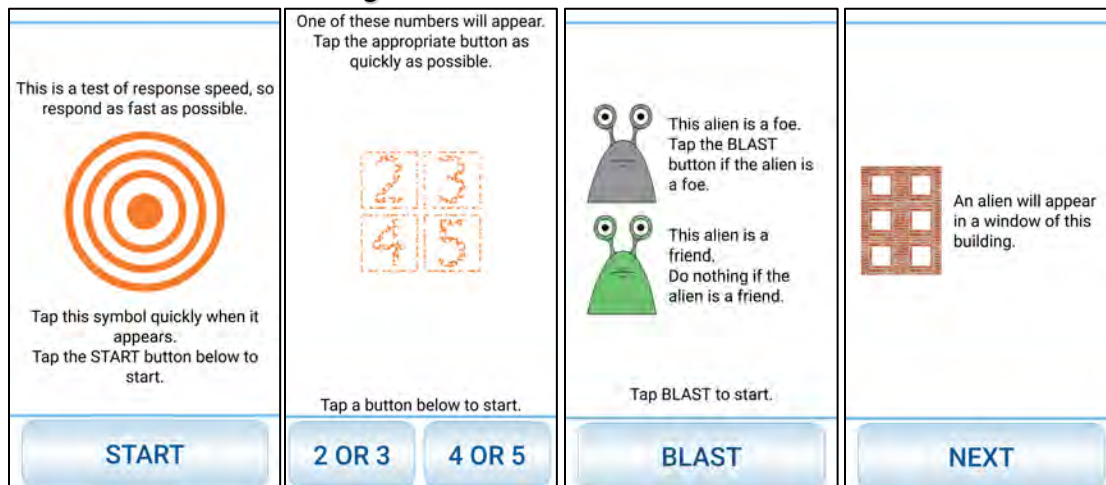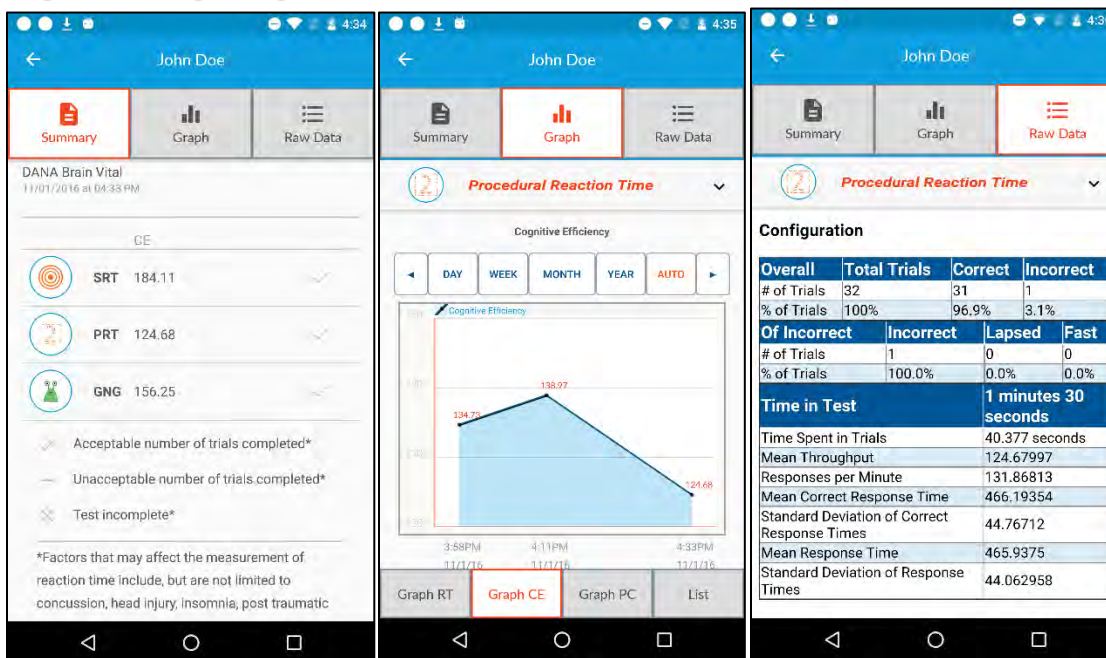

Figure 2: Sample Report Screens

II. Benefits of DANA

Because DANA runs as a mobile application on Android-based devices with results available immediately, Combat Medics, Navy Corpsmen, Air Force Pararescuemen (PJs), and other Role 1 providers can use it down range to assess warfighters in real-time; the Department of Defense (DoD) does not currently have this capability. DANA acts as a Brain Thermometer™ and provides fast and objective screening for any changes in cognitive and neurobehavioral function. It is extremely sensitive to changes in cognitive efficiency due to any cause, and provides a quick and simple means to capture longitudinal cognitive and neurobehavioral biometrics. It supports quicker triage and assessing fit-for-duty and return-to-duty determination. And, DANA tracks changes in cognitive processing over time, monitoring responsiveness to treatment.

Test performances are measured to the millisecond, and studies published in peer-reviewed journals show that test-retest reliability coefficient is higher than that of the Automated Neurobehavioral Assessment Metric (ANAM) for Simple Reaction Time (SRT) and Procedural Reaction Time (PRT) tests and of the Intermediate Post-Concussion Assessment and Cognitive Testing (ImPACT) for the SRT test (Russo and Lathan, 2015).

III. Risk Analysis

As Table 2 below shows, ATinc assesses that there is a low risk level associated with the technical, cost, schedule, and business areas related to DANA. In January 2017 ATinc successfully released a commercial version of DANA for the civilian market, so it is confident that the core technology performs per its specification. In addition, there have been multiple papers published in peer-reviewed scientific publications; citations for these articles are listed in Appendix A, establishing DANA's test-retest reliability and its ability to distinguish impaired vs. non-impaired in a variety of populations.

This is the first ATinc product to be integrated into the DoD's health care system. By leveraging long-term relationships with several DoD prime contractors, including Lockheed Martin, Raytheon BBN, among others, ATinc will ensure that it can integrate DANA successfully.

Table 2 Risk Chart

RISK AREA	RISK LEVEL		
	Low	Medium	High
BUSINESS	✓		
COST	✓		
SCHEDULE	✓		
TECHNICAL	✓		

IV.   Operational Needs

DANA addresses the DoD's need to objectively assess the cognitive processing of warfighters by Army Medics, Navy Corpsmen, Air Force PJs and other providers in Roles of Care 1-4, using a test that has a high test-retest reliability. In addition, DANA meets the DoD need for an FDA-cleared medical device, and since it runs as a mobile application on Android-based devices, it is suitable to be used at the point of injury, in Battalion Aid Stations, as well as at higher levels of care. DANA also provides the DoD with the ability to capture brain-based longitudinal biometrics over the entire length of service of a warfighter.

V.   Transition Targets

Special Operations Command (SOCOM). ATinc began actively working with SOCOM since August 2016 as a transition target. ATinc first met with the SOCOM Command Psychologist, COL Mark Baggett, USA, as well as SOCOM's Command Surgeon, CAPT Scott Cota, USN, at the Military Health System Research Symposium, held in Orlando, FL. At the meeting SOCOM identified DANA as potentially addressing its need to objectively and quickly assess Operators down range who may have experienced a change in cognitive processing. In addition, since DANA's test-retest reliability coefficient is higher than that of ImPACT for the SRT test, which SOCOM currently uses, it provides more statistically significant results. SOCOM is planning on conducting a pilot study using DANA in the summer of 2017, and if results from that study meet its expectations, is interested in acquiring DANA for use throughout the Command.

US Army Medical Research and Materiel Command (MRMC). Non-Invasive Neurocognitive Assessment Device (NINAD) Integrated Product Team (IPT). The NINAD IPT is the organization within the Army that is responsible for considering possible materiel solutions

for neurocognitive assessment devices. ATinc has been meeting with the NINAD IPT, and its members, on a regular basis since 2014. In those meetings, ATinc has updated the NINAD IPT on its progress in developing DANA, and most recently met with the NINAD IPT at an Industry Day held in Baltimore, MD in December 2016, and has answered follow-up questions from the IPT.

VI. TRL and MRL (Technology Readiness Level and Manufacturing Readiness Level)

TRL-7: System prototype demonstration in an operational environment (https://www.army.mil/e2/c/downloads/404585.pdf)

ATinc assesses DANA as TRL-7, since prototypes have been used in operational settings, including in Afghanistan, aboard the USS George Washington in high sea states, as well as in arctic, altitude, and jungle environments. Several articles have been published in peer-reviewed scientific journals, as listed in Appendix A, based on studies performed in those environments.

MRL-4: Capability to produce the technology in a laboratory environment (https://acc.dau.mil/CommunityBrowser.aspx?id=23209)

ATinc has successfully produced DANA for use by a number of DoD customers. However, these were very small production runs produced at ATinc's own facility. ATinc will need to scale its manufacturing capabilities, or identify a qualified subcontractor, when it comes time to produce DANA in production quantities.

VII. Technology Integration Process & Funding

DoD customers will be able to acquire and integrate DANA in at least two different ways:

1. Integration of DANA by DoD itself without support from ATinc

At the end of this Rapid Innovation Fund contract, ATinc will deliver to CDMRP on a DVD an executable copy for Android devices of DANA 4.0.0-RIF, and the DoD will have a royalty-free license to use this version of DANA. DoD customers within the DoD can then obtain DANA directly from CDMRP and integrate it into their medical protocols and Information Technology systems without any involvement from ATinc.

However, based on feedback that ATinc has received from potential DoD customers, such as SOCOM, it believes that they will need changes made to DANA to fit into their intended use cases, and so that it can be integrated with their medical Information Technology (IT) systems. For example, based on the SOW for this RIF contract, DANA needs to connect to a web portal and automatically upload data to that portal. We now know that will be unacceptable for most, if

not all, DoD customers. Under this scenario then, the DoD would need to modify DANA, without ATinc's assistance, to facilitate integration with the appropriate medical IT systems, such as with AHLTA-T or the Military Health System's GENESIS program.

Also, under this scenario, ATinc would not provide ongoing technical support or software updates to DANA. This could be problematic in the future as new versions of the Android OS are released and DANA may not be compatible with them. In addition, as noted in Section I above, ATinc currently tests Android devices to ensure that they are within specification for the variability in lag time in recording a user's response. If the DoD distributes and integrates DANA by itself, then it will also need to test and qualify new Android-based devices when they are released.

2. Integration of DANA with support from ATinc

Based on its experience in working with SOCOM, ATinc has a much better understanding of the work that needs to be done to successfully acquire and integrate DANA by a DoD customer, and would leverage that experience in working with other DoD customers. Below is the DANA technology integration plan that ATinc has developed for SOCOM based on its discussions with SOCOM, and then a more generalized DANA technology integration plan for other DoD customers.

A. DANA technology integration plan for SOCOM

Representatives from ATinc held a follow up meeting with COL Baggett and a number of members of his team at MacDill AFB in November 2016. As a result of that meeting, SOCOM purchased from ATinc one tablet and one mobile phone pre-loaded with DANA.

Based on feedback from SOCOM, ATinc agreed to make several changes to DANA, and completed DANA 4.1.0-SOCOM on March 31, 2017. These changes include the elimination of the automatic upload of data from DANA to ATinc's HIPAA compliant cloud and one button global export of data into CSV files. SOCOM intends to purchase a limited number of tablets with DANA from ATinc and begin a pilot study to collect data from warfighters participating in selection classes during the Summer, 2017. If that pilot study goes well, SOCOM has indicated to ATinc that it would then like to begin deploying DANA downrange.

In addition, SOCOM has informed ATinc that it would need the data from DANA to be exported into a separate database, SPEAR, which another contractor, Titus Human Performance Solutions, is providing to SOCOM. ATinc has worked with Titus to develop a proposal for SOCOM that would enable it to create a proof-of-concept integration – showing that DANA could in fact export data successfully into the SPEAR data base. ATinc is also

working with Titus to develop a proposal for SOCOM to support porting DANA so that it can run on the same Windows tablet as SPEAR currently runs.

In addition, from its discussions with SOCOM, ATinc has learned that a key feature that is included in the SOW under this RIF, automated connectivity to the ATinc cloud where DANA data would be automatically stored, is not acceptable to SOCOM, and probably other DoD customers, because of DoD-wide guidelines regarding the storage of medical data of active duty service members.

If SOCOM proceeds with acquiring DANA, ATinc will need to negotiate a contract to provide ongoing technical and software support for it. This contract would address the following areas:
    i. Creating training materials and user guides.
    ii. Developing a technical support plan so that ATinc can provide ongoing Level 1 and Level 2 technical support for DANA users.
    iii. Developing a software update plan so that ATinc can provide regular software updates for DANA that would, among other things, ensure its compatibility with future versions of the Android OS.
    iv. Testing and certifying Android smart phones and tablets to ensure DANA's reliability.

ATinc has developed the following notional plan for the integration of DANA into SOCOM:

Task	Date
Complete work on DANA 4.1.0-SOCOM, which includes changes to address SOCOM-specific needs.	March 31, 2017
Load DANA 4.1.0-SOCOM onto Android OS tablet and smart phone already purchased by SOCOM.	April 15, 2017
Support SOCOM Trial/data collection: • Receive and process order for 50 tablets pre-loaded with DANA. • Support SOCOM trial and data collection itself.	April 15-September 30, 2017
Proof-of-Concept migration of data from DANA to SPEAR database.	July 1-September 30, 2017
Follow up to SOCOM trial: • Determine with SOCOM any needed revisions to DANA based on experience from trial.	October 1, 2017-March 30, 2018

• Implement and test needed revisions. • Deliver revised DANA.	
Initial roll out of DANA to SOCOM.	April 1, 2018

## B. DANA technology integration plan for other DoD customers

Listed below are the areas which ATinc would address, under acquisition and service contracts to be negotiated, with other DoD customers to ensure the successful technology integration of DANA:

a. Modify DANA 4.0.0-RIF to address the needs of specific DoD customers and their intended use cases.
b. Develop a healthcare IT integration plan, which could entail:
   (1) Creation of a Risk Management Framework (RMF) for integration of DANA Into DoD healthcare IT systems.
   (2) Process to export data into AHLTA-T or other healthcare IT systems, such as MHS GENESIS, to be determined based on customer needs.
   (3) Ensure compatibility with requirements of appropriate Program Offices such as the Joint Operational Medicine Information Systems (JOMIS) Program Management Office (PMO) or the MC 4 Program Office. The JOMIS PMO, for example, is in the process of introducing the Mobile Computing Capability, which utilizes Android-based devices to host medical mobile applications, which could be devices on which DANA runs.
c. Distribute DANA in the following ways:
   (1) Create and maintain an FTP site from which authorized DoD customers could download DANA.
   (2) Provide DANA preloaded on Android OS smart phones and tablets.
      (a) If this option was selected, approval of a Manufacturing Plan to support the acquisition of Android devices and loading of DANA onto them at scale.
d. Update training materials and user guides as appropriate for any changes made to DANA in response to the DoD's needs.
e. Develop a technical support plan so that ATinc can provide ongoing Level 1 and Level 2 technical support for DANA users.
f. Develop a software update plan so that ATinc can provide regular software updates for DANA that would, among other things, ensure its compatibility with future versions of the Android OS.
g. Test and certify Android smart phones and tablets to ensure DANA's reliability.

Appendix A

Peer-Reviewed Publications Utilizing DANA

1. Hollinger, K. R., Franke, C., Arenivas, A., Woods, S. R., Mealy, M. A., Levy, M., & Kaplin, A. I. (2016). Cognition, mood, and purpose in life in neuromyelitis optica spectrum disorder. Journal of the neurological sciences, 362, 85-90.

2. Lathan, Corinna, et al. "Defense Automated Neurobehavioral Assessment (DANA)- psychometric properties of a new field-deployable neurocognitive assessment tool." Military medicine 178.4 (2013): 365-371.

3. Lathan CE, Coffman I, Shewbridge R, Lee M, Cirio R, et al. A Pilot to Investigate the Feasibility of Mobile Cognitive Assessment of elderly patients and caregivers in the home. J Geriatrics Palliative Care 2016;4(1): 6.

4. Roach, Emma B., et al. "AltitudeOmics: Decreased reaction time after high altitude cognitive testing is a sensitive metric of hypoxic impairment."NeuroReport 25.11 (2014): 814.

5. Russo, C. R., and C. E. Lathan. "An Evaluation of the Consistency and Reliability of the Defense Automated Neurocognitive Assessment Tool." Applied Psychological Measurement (2015): 0146621615577361

6. Spira, James L., et al. "The Impact of Multiple Concussions on Emotional Distress, Post-Concussive Symptoms, and Neurocognitive Functioning in Active Duty United States Marines Independent of Combat Exposure or Emotional Distress." Journal of neurotrauma 31.22 (2014): 1823-1834.

7. Subudhi, Andrew W., et al. "AltitudeOmics: the integrative physiology of human acclimatization to hypobaric hypoxia and its retention upon reascent." PloS one9.3 (2014): e92191.
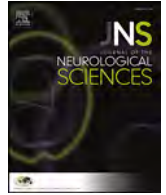
Cognition, mood, and purpose in life in neuromyelitis optica spectrum disorder

Kristen R. Hollinger [a,b], Caroline Franke [a], Ana Arenivas [a,c], Steven R. Woods [a], Maureen A. Mealy [b], Michael Levy [b], Adam I. Kaplin [a,b,*]

[a] Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, United States
[b] Department of Neurology, Johns Hopkins University, Baltimore, MD, United States
[c] Department of Neuropsychology, Kennedy Krieger Institute, Baltimore, MD, United States

1. Introduction

Neuromyelitis optica spectrum disorder (NMOSD) is a rare neurological condition that affects an estimated 4000–8000 people in the United States, making it about 100 times less prevalent than multiple sclerosis (MS) [1]. The criteria for a diagnosis of NMOSD, newly developed by the International Panel for NMOSD diagnosis, requires either the presentation of at least one core clinical characteristic of NMOSD and aquaporin-4 (AQP4)-IgG positivity in the absence of alternative diagnoses, or the presentation of at least two disseminated core clinical characteristics (one being optic neuritis, acute myelitis with longitudinally extensive transverse myelitis lesions, or area postrema syndrome) and fulfillments of MRI requirements in the case of optic neuritis acute myelitis, area postrema syndrome, and acute brainstem syndrome if the patient has a negative or unknown AQP4-IgG status [2].

Until recently it was assumed that the brain is largely spared in NMOSD, which led to the conclusion that higher cortical functions also remain unaffected by the disease. However, more recent studies report highly selective brain injury in NMOSD. The regions and extent of damage vary depending on the report, with some reporting damage to white matter but no loss of gray matter [3], while others report that white matter is spared but normal appearing gray matter is compromised in NMOSD [4]. Others account damage to both white and gray matter in NMOSD [5].

Cognition has only recently been studied in NMOSD, with the first report of impairments in cognitive function in the disease published in 2008 [6]. In two limited studies on cognition in NMOSD, the rates of cognitive impairment in NMOSD are similar to those observed in patients with MS, with approximately half of all NMOSD and MS patients displaying impaired cognitive performance [3,7]. Significant impairments in speed of information processing and sustained attention occur in NMOSD, and the degree of impairment is also similar to that observed in MS [6].

Few studies have characterized features of depression in NMOSD, with the first case of depression related to NMOSD reported in 2004 [8]. Similar to MS [9,10], rates and severity of depression are higher in NMOSD as compared to the general population [11]. In fact, two studies comparing depressive symptoms in NMOSD and MS reported more severe depression in NMOSD versus MS [12,13]. Depression has been linked to cognition in NMOSD, with more severe depression associated with worse cognitive function, while disease duration and EDSS scores do not appear to be related to mood in NMOSD [13].

The Purpose in Life (PIL) test was developed by Crumbaugh and Maholick based on the teaching of Dr. Viktor Frankl [14]. Dr. Frankl survived life in a concentration camp during the Holocaust, and during this time he noticed that those who seemed to have a higher purpose in life were more likely to survive. The PIL covers three dimensions of life purpose: the will to find meaning in existence, the freedom to create meaning in daily activities, and the will to find meaning in future challenges [15]. Higher life purpose has been linked to decreased incidence of a variety of physical ailments, including Alzheimer's disease (AD) [16], incident disability in the elderly [17], stroke [18], and myocardial infarction [19,20]. High purpose in life is also associated with increased participation in physical activity [21], not smoking [22], and use of

* Corresponding author at: Johns Hopkins Department of Psychiatry, 600 N. Wolfe St, Meyer 1-121, Baltimore, MD 21287, United States.
E-mail address: akaplin@jhmi.edu (A.I. Kaplin).

Table 1
Summary of DANA cognitive tests.

Cognitive test	Structure	Targets
Code Substitution	9 symbol-digit pairs are shown in a key and one combination is displayed, subject determines if the combination matches the key.	Executive capacity, immediate memory, and attention
Spatial Processing	Pairs of 4-bar histograms are presented, one rotated 90°, subject determines if they are the same or different	Executive capacity and spatial manipulation
Simple Reaction Time	Subject taps on the location of an asterisk symbol as soon as it appears	Reaction time
Procedural Reaction Time	One of 4 numbers is displayed, subject must select if the number is a "2 or 3", or "4 or 5"	Executive functioning with decision making capabilities
Finger tapping test	Subject taps the screen with pointer finger of dominant hand as many times as possible within a given time	Motor function

preventative health care services [23]. PIL in NMOSD has not yet been studied.

Here, we sought to measure the relationship between cognition, mood, and PIL in NMOSD. This is the first study to assess PIL in NMOSD. Here, we compare PIL in NMOSD to non-NMOSD control subjects, and we examine the relationships between PIL, mood, and cognition in both cohorts.

2. Material and methods

2.1. Participants

Subjects were recruited from attendees of the Johns Hopkins Hospital NMO Patient Day, held on October 5, 2014. Attendees came from across the United States to attend a series of lectures, and numerous research studies were conducted in conjunction with the lectures. Family or friends of NMOSD patients also attending NMO Patient Day served as control subjects. Only willing participants were recruited into the study, and those attending NMO Patient Day who did not wish to participate in research studies could do so without penalty. 20 control subjects and 23 individuals with NMOSD completed a DANA battery of cognitive assessment tests (see description below), with an additional 1 NMOSD patient completing the Patient Health Questionnaire (PHQ-9) test and the PIL test, and another NMOSD patient completing the PIL test (total NMOSD n = 25). All participants provided general personal information, including age, gender, and highest level of education. NMOSD participants provided information related to their disease, including date of NMOSD onset, date of NMOSD diagnosis, number of relapses, time since last relapse, and current mobility status. Three of the 25 NMOSD subjects are AQP4-IgG seronegative, but all 25 meet the diagnostic criteria for NMOSD. All protocols were approved by the Johns Hopkins Institutional Review Board.

2.2. DANA cognitive assessment battery

Study subjects underwent a battery of cognitive tests on the neurocognitive assessment tool Defense Automated Neurobehavioral Assessment (DANA), developed by AnthroTronix, Inc. (Silver Spring, MD) [24]. The DANA tests are conducted on Samsung Galaxy tablets. The cognitive tests included in the current study were the Simple Reaction Time (SRT) test, Procedural Reaction Time (PRT) test, Spatial Processing (SP), and Code Substitution (CS) (Table 1). The primary outcome for cognitive tests was throughput, calculated as [(% correct) / (Reaction Time for correct responses) × 60,000]. In addition to the cognitive tests, the DANA battery also included a finger tapping test (FTT) to assess tapping motor function and the PHQ-9 to assess mood. In the FTT, the patient taps the tablet screen as many times as he or she can in a 10-second interval. Three consecutive trials were conducted with the dominant hand used to complete the cognitive test battery. The PHQ-9 is a standard and valid 9-question test to evaluate depression severity based on symptoms within the last two weeks [25].

2.3. Purpose in life

A subset of participants additionally completed a modified Purpose in Life (PIL) survey. The PIL survey consists of 20 Likert-style items, in which the patient self-ranks himself or herself on a scale of 1–7 with anchors to each question or statement [14]. For example, question 1 reads, "I am usually: 1(completely bored), 2, 3, 4 (neutral), 5, 6, 7 (exuberant, enthusiastic)", and the subject circles the number corresponding to his or her usual state.

2.4. Statistical analyses

Regression analyses were conducted using Stata 13.1 (College Station, TX). Univariate analyses were conducted for DANA cognitive tests, followed by multivariate analyses. Independent variables factored into the multivariate analyses included age, gender, highest level of education, mood (as determined by PHQ-9 test), and the reported number of hours of sleep the previous night. Data are presented as Mean ± SEM. P values less than 0.05 are considered statistically significant.

3. Results

Study participant information is presented in Table 2. The control group had a nearly equal ratio of males and females. Females were significantly overrepresented in the NMOSD group (n = 23/25, 92%, P < 0.01), in line with the 6.5:1 female predominance of NMOSD in American patients [26]. Mean age of participants was higher in controls versus NMOSD (50.97 ± 3.48 versus 44.03 ± 2.86 years, respectively), but this difference did not reach statistical significance (P = 0.13). Participants attained equivalent levels of education (Control = 14.7 ± 0.6, NMOSD = 14.96 ± 0.5).

DANA cognitive test responses require a finger tap from the participant. Because NMOSD subjects can have motor impairments that might

Table 2
Study Sample Characteristics.

	NMOSD n = 25	Control n = 20	P value
Gender (female)	23 (92%)	11 (55%)	0.004
Mean age (years)	44.03 ± 2.86	50.97 ± 3.48	0.128
Level of education (years)	14.96 ± 0.50	14.70 ± 0.59	0.736
Hours sleep, previous night	5.89 ± 0.39	6.58 ± 0.29	0.173
Disease duration (months)	65.09 ± 11.84		
# of days since last relapse	413.6 ± 131.0		
Total # of relapses	4.25 ± 0.82		
Mobility impairment			
Fully mobile	14 (56%)		
Occasional walking aid (e.g. cane)	3 (12%)		
Walking aid required	1 (4%)		
Occasional wheelchair	3 (12%)		
Wheelchair-bound	4 (16%)		

Data are presented as mean ± SEM or number of participants (% total).

confound cognitive tests results, FTTs were conducted to determine if differences in finger motor function existed between groups. Interestingly, nearly statistically significant differences were observed between groups in the first FTT test, with NMOSD participants displaying fewer finger taps than controls (53.86 ± 2.09 versus 58.95 ± 1.47, respectively, p = 0.058), but the differences disappeared by the second (54.23 ± 2.14 versus 56.20 ± 1.87, respectively, P = 0.496) and third trials (52.18 ± 1.73 versus 51.80 ± 2.72, respectively, P = 0.905). Because the FTT results normalized between groups by the second and third trial and FTT tests were conducted prior to the cognitive tests, FTT results were not factored into primary analyses of cognitive test data.

No differences were observed between groups in uncontrolled univariate analyses for any cognitive test (SRT, PRT, SP, and CS; P = 0.46–0.95). When controlled for individually in multivariate analyses, gender, highest level of education, PHQ-9 score, and number of hours of sleep did not impact this significance of these results (P = 0.27–0.77). However, when age was controlled for in multivariate analyses, statistical significance was reached between NMOSD and control performance in the Code Substitution (CS) test (P = 0.037). Significance increased further when gender, highest level of education, PHQ-9 score, and number of hours of sleep were controlled for in addition to age in multivariate analysis (P = 0.029). Specifically, NMOSD patients had a 17.8% decrease in throughput scores compared to control subjects, indicating cognitive impairment in NMOSD. Similarly, NMOSD patients had lower mean correct response time on the CS test, as they took 13.8% longer to select the correct response versus controls (P = 0.022).

In the NMOSD group, neither disease duration nor diagnosis duration correlated with cognitive performance on any test. Significant relationships were observed, however, between the throughput metric on all four cognitive tests and age (Table 3). Significant associations between age and CS throughput were observed in all subjects pooled (P < 0.001), with every year of age corresponding to a 0.325 drop in CS throughput (mean control CS throughput = 35.02). The significant age and CS throughput relationships persisted after separating out control subjects (P < 0.001), and the trend remained after separating out NMOSD subjects (P = 0.074). Similar patterns emerged in the Spatial Processing (SP) test, with significant negative correlations existing in pooled and grouped analyses (P < 0.05 for all analyses). Negative relationships between age and throughput on the PRT and SRT tests were observed in pooled and control analyses (P < 0.05 for all analyses), but associations were not observed between age and throughput in the NMOSD group.

There was no difference in the number of hours of sleep the previous night of testing between groups. A significant relationship between CS throughput and the reported number of hours of sleep the previous night was observed in NMOSD patients, with every hour of sleep corresponding with a 1.88 increase in CS throughput performance (P = 0.03). The reported number of hours of sleep the previous night did not affect CS throughput score in control subjects (P = 0.584).

Scoring of the PHQ-9 is based on five point increments, where 0–4 indicates no depression, 5–9 indicates mild depression, 10–14 indicates moderate depression, 15–19 indicates moderately severe depression, and 20–27 indicates severe depression [27]. 58.3% of NMOSD subjects displayed mild, moderate or moderately severe depression compared

Table 3
Age versus cognitive test throughput regression analyses.

Cognitive test	All subjects		Control		NMOSD	
	P value	coefficient	P value	coefficient	P value	coefficient
Code Substitution	<0.001	−0.325	<0.001	−0.499	0.074	−0.218
Spatial Processing	<0.001	−0.22	0.001	−0.278	0.025	−0.185
Simple Reaction Time	0.024	−0.716	0.014	−0.995	0.224	−0.662
Procedural Reaction Time	0.006	−0.368	0.001	−0.599	0.595	−0.111

to 21.1% of control subjects. 37.5% (9/24) NMOSD patients displayed mild depression, 16.7% (4/24) displayed moderate depression, and 4.2% (1/24) displayed moderately severe depression as compared to 5.3% (1/19), 5.3% (1/19), and 10.5% (2/19) of control subjects with mild, moderate, and moderately severe depression, respectively. Average PHQ-9 scores did not differ significantly between groups, although NMOSD patients did have higher scores, indicating more depressed mood overall, versus the control group. Control subjects averaged 4.26 ± 1.32 on the PHQ-9 test, corresponding to no depression, while NMOSD patients averaged 6.38 ± 0.95 points, corresponding to mild depression. In pooled analysis, females had higher depression scores compared to males (P = 0.01), and significance was maintained after controlling for diagnosis of NMOSD (P = 0.03).

PIL data revealed no differences between average PIL score in control and NMOSD subjects (113.5 ± 3.1 vs. 109.5 ± 2.7). There were, however, opposing relationships between cognition and PIL in control and NMOSD groups. CS throughput score improved by 0.205 points for every one point increase in PIL score in the NMOSD group (P = 0.067), while CS throughput score went down by 0.335 points for every one point increase in PIL score in the control group (P = 0.039) (Fig. 1). A trend relation was observed between PIL and PHQ-9 scores in control populations (P = 0.088), where higher PIL trended with lower PHQ-9 scores (i.e. less depression) in both groups. No relationship was found between PHQ-9 score and PIL in NMOSD subjects. The relationship between PIL and mood became statistically significant, however, when groups were pooled together (P = 0.041, Fig. 2).

4. Discussion

The present study is the first to evaluate and compare cognition, mood, and PIL in NMOSD and control subjects. We found significant impairments on CS test performance in NMOSD subjects after controlling for age, mood, education, gender, and number of hours of sleep the previous night. No impairment of cognitive function in NMOSD subjects was observed, however, when personal and demographic information were not taken into account in analyses. To study the effect of each variable (age, mood, education, gender, and number of hours of sleep), additional analyses were run in which each variable was omitted. Individual omission of gender, mood, education, or number of hours of sleep the previous night caused only minor fluctuations in significance level of CS throughput between groups (0.016–0.056). Omission of age, however, completely erased significant differences between groups (P = 0.44), indicating that cognitive test performance was highly dependent on age. In line with this observation, highly significant correlations were detected in pooled analyses between age and throughput in all cognitive tests, with test performance dropping as age increases. There was a trend toward differences between the average ages of NMOSD and control groups that did not reach statistical significance, with control subjects older than NMOSD subjects by over 6 years. Therefore, the age difference between groups would have concealed cognitive impairment in NMOSD if not properly controlled.

The CS test employed in the present study is similar to the more popular symbol digit modalities test (SDMT). In both the CS and SDMT test, 9 pairs of numbers and symbols are presented in a key, and the participant must communicate the correct number that corresponds to presented symbols. The SDMT is one of the most widely used cognitive tests in MS to rate cognitive impairment, due to its ease of use, high test–retest reliability, and high sensitivity [28,29]. A study in Chinese NMO patients reported impairments in the SDMT test (P < 0.001) as compared to age- and gender-matched controls [30]. This finding, in agreement with our CS test results, indicates that both the CS and SDMT are sensitive to detect cognitive impairment in NMOSD.

It is interesting that the only test to detect cognitive differences in NMOSD and control subjects in the present study was the CS test. A possible explanation as to why the SRT, PRT, and SP tests did not show impairments in NMOSD patients is because our patient
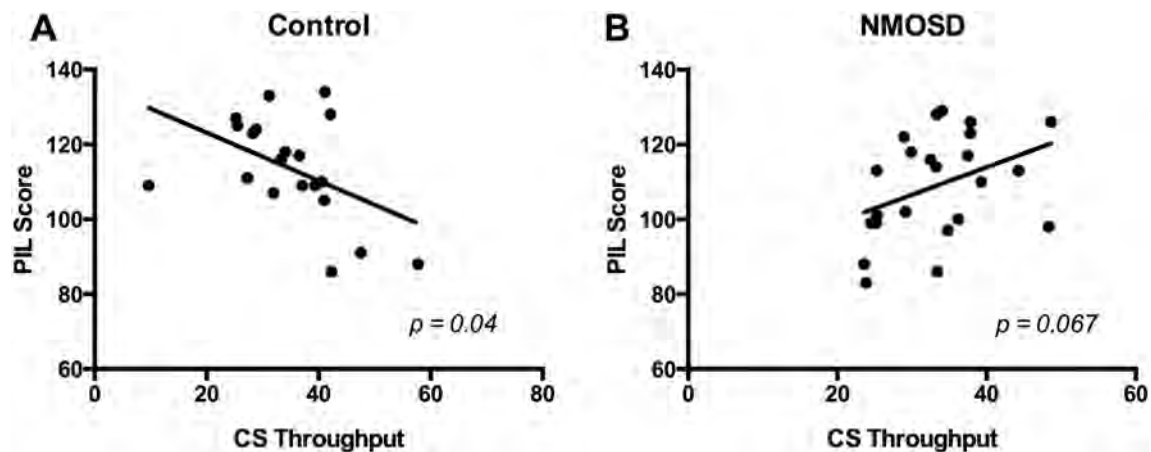
Fig. 1. Higher Purpose in Life (PIL) scores are associated with poor cognition as measured by the Code Substitution (CS) test in control subjects (A). Conversely, PIL and CS test performance are positively associated in NMOSD subjects (B).

population was highly educated (14.96 ± 0.5 years). Lower education levels are predictive of cognitive impairment in NMOSD [30]. Most NMOSD studies typically examine patient cohorts with an average of 11–12 years of education [3,6,13]. It is therefore possible that cognitive impairment was only detected in the CS test in the present study because our NMOSD patient population had such high levels of education attainment and were therefore more protected against cognitive impairment than typical NMOSD patient populations. The CS test could be the only test sensitive enough to pick up cognitive impairment in our highly educated patient population. The nature of the experimental design in which we tested NMOSD patients who willingly attended NMO Patient Day could select for more highly educated participants. Future studies will be designed to include a more diverse and representative NMOSD patient cohort. Within the NMOSD patient population, we also did not observe age-depended decreases in performance in the SRT and PRT tests. These two tests are simpler and require less cognitive processing than the SP and CS tests, so it is possible that physical aspects of NMOSD cloud age-dependent changes in performance.

Disease duration, mobility, and sleep were examined to determine if any of these factors were related to cognition in NMOSD. Similar to other reports [7], we did not observe a relationship between NMOSD disease duration and cognitive test performance, suggesting that cognitive impairment occurs early in some individuals, but for others it is not an inevitable and progressive comorbidity of the disease. The present study required a definite diagnosis of NMO, but others have shown that diagnosis of NMO subtype (limited NMO versus definite NMO) does not relate to cognition [7]. In fact, cognitive impairment can be present in NMO patients without any visible brain lesions [31].

In the present study, 44% (11/25) of NMOSD participants reported some mobility impairment, but severity of mobility impairment also did not correlate with any cognitive test result or upper extremity dexterity (as assessed by FTT results). There was, however, a relationship between cognitive test performance and sleep in NMOSD patients, with NMOSD patients who had significantly less sleep the previous night performing worse on cognitive tests. This relationship between hours of sleep and cognition did not exist in the control subjects, suggesting that NMOSD patients may be more vulnerable to the cognitively impairing effects of little sleep. The importance of a good night's sleep has been demonstrated in many studies in healthy individuals [32], but these data suggest the importance of sleep in NMOSD patients.

Use of the PHQ-9 to screen for depression has been validated in patients with multiple sclerosis (MS) [33]. Average PHQ-9 scores for non-NMO controls placed them in the non-depressed category, while
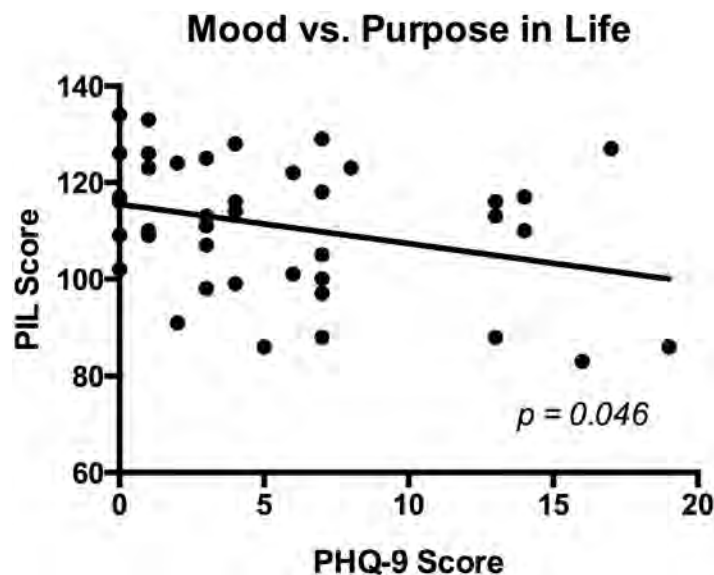

Fig. 2. Pooled data from control and NMOSD participants reveal a significant relationship between mood and PIL. Lower PHQ-9 scores, indicating less signs of depression, are associated with higher scores in the PIL test, indicating a greater sense of purpose in life.

the average PHQ-9 score of the NMOSD group fell into the mildly depressed category. Although it would be reasonable to hypothesize that level of disability would impact mood, with more disabled patients having lower mood, PHQ-9 scores were unrelated to mobility status in NMOSD subjects. Similar results have been obtained in MS patients [34,35], indicating that physical disability alone does not cause depression in these neurological diseases. Similar to other reports [11], we found higher rates of depression in NMOSD versus control subjects. Although sleep disturbances are common in depressed individuals [36], we did not observe a relationship between number of hours of sleep the previous night and depression scores in either control or NMOSD subjects.

An interesting observation of the present study was the conflicting relationships between cognition and PIL between the NMOSD and control groups. While no differences in average PIL existed between groups, control subjects with a higher PIL score did worse on the CS cognitive test, while NMOSD subjects with a higher PIL score did better on the CS cognitive test. The results of this NMOSD study are in line with a study in AD patients that showed high PIL protects against cognitive impairment in AD when neuropathological burden is high [37]. It is therefore possible that high PIL could be protective against cognitive impairment in NMOSD as well. No studies have reported on PIL in primary caregivers of patients with physical disability. While not all control subjects in the present study were primary caregivers, all had a close friendship or familial link to the NMOSD subjects as significant travel was required for most study participants. It is possible that the stress of caregiver burden could negatively impact PIL or cognition in the control cohort. Future studies will include the differentiation between non-NMO controls who are not caregivers and controls who are caregivers to examine the effect of NMOSD caregiver burden on PIL and cognition.

A recent study evaluated computerized touch screen testing in subjects with NMO and MS [38]. Similar domains of cognition were tested in their cognitive battery and the present study. Although the authors recognized that cognitive impairment is present in NMO, they did not detect differences between the cognitive performance of 10 NMO patients and 15 control subjects as measured by the computerized tests. Based on these results, the authors concluded that computerized touch testing is not useful or sensitive in NMO patients. It is interesting, however, that no differences were observed in the mini-mental state examination (MMSE) scores between control subjects and participants with NMO. The MMSE is a standard test of cognitive function that has been used for over 40 years in research studies. With both the MMSE and the computerized test data in mind, it appears that a computerized test battery is relevant for measuring cognition in NMO, but that their sample size was too small to detect cognitive impairment in NMO by any measure. A further benefit of computerized testing, like the DANA test employed in the present study, is increased sensitivity. There is a maximum of 30 possible points to earn on the MMSE, so a ceiling effect is easily reached in certain cohorts. The DANA test, however, times tests down to the millisecond, allowing for extremely sensitive results that span a wider range between subjects.

There are several limitations to the present study. The participant group is self-selected on two dimensions. First, participants had to be interested in and motivated to attend Johns Hopkins NMO Patient Day. Therefore, it is possible that the attendees had better mood, higher PIL, and/or superior cognition as compared to the general NMOSD population. Second, attendance required significant travel for many of the participants, which could thwart the participation of some individuals with severe physical impairments and those in a lower socioeconomic status. While any of these factors could affect our results, the inclusion of family and friends of the NMOSD patients as the control population likely controlled for possible differences in socioeconomic status and/or stress of travel to attend Patient Day, and the variables of age, gender, education history, mood, and sleep were controlled for in the statistical analyses. As discussed above, however, rates of cognitive impairment could be lower in the present study as it is possible that high levels of

education protect against cognitive impairment in NMOSD. Another limitation of the present study is that pain was not included in the analyses. While no subjects were in active or visible pain at any time during the testing process, it is possible that pain could influence our findings as others have demonstrated potential relationships between pain and depression [12], and pain and cognitive impairment in NMOSD [7]. Future studies will factor pain into analyses to determine if there is a relationship between pain and mood, cognition, and/or PIL. A final limitation of the present study is the lack of MRI data. We do not treat all participants with NMOSD who attended the Johns Hopkins NMO Patient Day, so MRI data were not available to us from all research subjects. Few studies have measured the relationship between cognition and brain lesions in NMOSD, but a recent report suggests that gray but not white matter lesions are associated with cognitive impairment in the disease [39]. Although outside the scope of the present study, future work from our group will focus on functional connectivity and white matter integrity in NMOSD.

5. Conclusions

Taken together, our results support the use of a mobile tablet-based cognitive assessment tool utilizing CS, an electronic version of the SDMT, for measuring cognitive impairment in NMOSD patients. Depression and cognitive impairment are comorbidities of NMOSD that are only recently beginning to be understood. Further studies are required to better characterize the interplay of mood and cognition in NMO. NMOSD patients with high PIL perform better on cognitive tests, and future studies will be designed to ascertain whether a higher will to find meaning in existence is a protective factor against cognitive decline in NMOSD.

Conflict of interest

The authors have no conflicts of interest to disclose.

Acknowledgments

References

[1] M.A. Mealy, et al., Epidemiology of neuromyelitis optica in the United States: a multicenter analysis, Arch. Neurol. 69 (9) (2012) 1176–1180.
[2] D.M. Wingerchuk, et al., International consensus diagnostic criteria for neuromyelitis optica spectrum disorders, Neurology (2015).
[3] F. Blanc, et al., White matter atrophy and cognitive dysfunctions in neuromyelitis optica, PLoS ONE 7 (4) (2012) e33878.
[4] M.A. Rocca, et al., Magnetization transfer and diffusion tensor MRI show gray matter damage in neuromyelitis optica, Neurology 62 (3) (2004) 476–478.
[5] C.S. Yu, et al., Diffusion tensor imaging in the assessment of normal-appearing brain tissue damage in relapsing neuromyelitis optica, AJNR Am. J. Neuroradiol. 27 (5) (2006) 1009–1015.
[6] F. Blanc, et al., Cognitive functions in neuromyelitis optica, Arch. Neurol. 65 (1) (2008) 84–88.
[7] E. Saji, et al., Cognitive impairment and cortical degeneration in neuromyelitis optica, Ann. Neurol. 73 (1) (2013) 65–76.
[8] A. Borden, et al., Depression and Devic's syndrome, Am. J. Psychiatry 161 (6) (2004) 1128.
[9] R.A. Marrie, et al., The incidence and prevalence of psychiatric disorders in multiple sclerosis: a systematic review, Mult. Scler. 21 (3) (2015) 305–317.
[10] S. Mrabet, et al., Depression in multiple sclerosis, Rev. Neurol. (Paris) 170 (11) (2014) 700–702.
[11] D. He, et al., Cognitive function, depression, fatigue, and activities of daily living in patients with neuromyelitis optica after acute relapse, Int. J. Neurosci. 121 (12) (2011) 677–683.
[12] P. Qian, et al., Association of neuromyelitis optica with severe and intractable pain, Arch. Neurol. 69 (11) (2012) 1482–1487.

[13] S. Vanotti, et al., Cognitive performance of neuromyelitis optica patients: comparison with multiple sclerosis, Arq. Neuropsiquiatr. 71 (6) (2013) 357–361.
[14] J.C. Crumbaugh, L.T. Maholick, An experimental study in existentialism: the psychometric approach to Frankl's concept of noogenic neurosis, J. Clin. Psychol. 20 (1964) 200–207.
[15] E. Jonsen, et al., Psychometric properties of the Swedish version of the Purpose in Life scale, Scand. J. Caring Sci. 24 (1) (2010) 41–48.
[16] P.A. Boyle, et al., Effect of a purpose in life on risk of incident Alzheimer disease and mild cognitive impairment in community-dwelling older persons, Arch. Gen. Psychiatry 67 (3) (2010) 304–310.
[17] P.A. Boyle, A.S. Buchman, D.A. Bennett, Purpose in life is associated with a reduced risk of incident disability among community-dwelling older persons, Am. J. Geriatr. Psychiatr. 18 (12) (2010) 1093–1102.
[18] E.S. Kim, et al., Purpose in life and reduced incidence of stroke in older adults: 'The Health and Retirement Study', J. Psychosom. Res. 74 (5) (2013) 427–432.
[19] L. Yu, et al., Purpose in life and cerebral infarcts in community-dwelling older people, Stroke 46 (4) (2015) 1071–1076.
[20] E.S. Kim, et al., Purpose in life and reduced risk of myocardial infarction among older U.S. adults with coronary heart disease: a two-year follow-up, J. Behav. Med. 36 (2) (2013) 124–133.
[21] S.A. Hooker, K.S. Masters, Purpose in life is associated with physical activity measured by accelerometer, J. Health Psychol. (2014).
[22] B. Konkoly Thege, et al., Relationship between meaning in life and smoking status: results of a national representative survey, Addict. Behav. 34 (1) (2009) 117–120.
[23] E.S. Kim, V.J. Strecher, C.D. Ryff, Purpose in life and use of preventive health care services, Proc. Natl. Acad. Sci. U. S. A. 111 (46) (2014) 16331–16336.
[24] C. Lathan, et al., Defense Automated Neurobehavioral Assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool, Mil. Med. 178 (4) (2013) 365–371.
[25] R.D. Kocalevent, A. Hinz, E. Brahler, Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population, Gen. Hosp. Psychiatry 35 (5) (2013) 551–555.
[26] W.L. Pereira, et al., Epidemiological, clinical, and immunological characteristics of neuromyelitis optica: a review, J. Neurol. Sci. (2015).
[27] M. Zimmerman, Symptom severity and guideline-based treatment recommendations for depressed patients: implications of DSM-5's potential recommendation

[28] R.H. Benedict, et al., Reliability and equivalence of alternate forms for the Symbol Digit Modalities Test: implications for multiple sclerosis clinical trials, Mult. Scler. 18 (9) (2012) 1320–1325.
[29] J. Van Schependom, et al., The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis, Eur. J. Neurol. 21 (9) (2014) 1219–1225 e71-2.
[30] N. Zhang, et al., Cognitive impairment in Chinese neuromyelitis Ooptica, Mult Scler (2015).
[31] D. He, et al., Cognitive impairment and whole brain diffusion in patients with neuromyelitis optica after acute relapse, Brain Cogn. 77 (1) (2011) 80–88.
[32] M.K. Scullin, D.L. Bliwise, Sleep, cognition, and normal aging: integrating a half century of multidisciplinary research, Perspect. Psychol. Sci. 10 (1) (2015) 97–137.
[33] K. Sjonnesen, et al., Evaluation of the 9-item Patient Health Questionnaire (PHQ-9) as an assessment instrument for symptoms of depression in patients with multiple sclerosis, Postgrad. Med. 124 (5) (2012) 69–77.
[34] C. Chruzander, et al., A 10-year follow-up of a population-based study of people with multiple sclerosis in Stockholm, Sweden: changes in disability and the value of different factors in predicting disability and mortality, J. Neurol. Sci. 332 (1–2) (2013) 121–127.
[35] J. Nielsen-Prohl, et al., Stress-stimulated volitional coping competencies and depression in multiple sclerosis, J. Psychosom. Res. 74 (3) (2013) 221–226.
[36] A.B. Medina, et al., Update of sleep alterations in depression, Sleep Sci. 7 (3) (2014) 165–169.
[37] P.A. Boyle, et al., Effect of purpose in life on the relation between Alzheimer disease pathologic changes on cognitive function in advanced age, Arch. Gen. Psychiatry 69 (5) (2012) 499–505.
[38] Y. Kawahara, et al., Cognitive and affective assessments of multiple sclerosis (MS) and neuromyelitis optica (NMO) patients utilizing computerized touch panel-type screening tests, Intern. Med. 53 (20) (2014) 2281–2290.
[39] Y. Liu, et al., Structural MRI substrates of cognitive impairment in neuromyelitis optica, Neurology 85 (17) (2015) 1491–1499.

of the PHQ-9 as the measure of choice for depression severity, Psychother. Psychosom. 81 (6) (2012) 329–332.

A Pilot to Investigate the Feasibility of Mobile Cognitive Assessment of elderly patients and caregivers in the home

Abstract

Background: The number of older adults with Alzheimer's disease (AD) has been steadily increasing and is likely to triple by 2050. Parallel increases in AD and informal AD caregivers who experience their own physical and cognitive challenges will result in the need for tools that can help both populations track their cognitive health easily, both in the clinic and at home.

Methods: DANA, a tablet-based, FDA-cleared computerized cognitive assessment tool, was used over 90 days among seven caregiver-AD patient dyads in-clinic and at home to assess DANA's sensitivity in detecting mild cognitive impairment and dementia as well as its feasibility in the home and clinic.

Results: DANA is sensitive to certain differences in cognitive performance between AD patients and caregiver. Most subtests were found to be feasible for in-home use among both patients and caregivers.

Conclusion: DANA shows promise for use both in-clinic and in the home to track cognitive performance of AD patients and their caregivers.

Introduction

About one in nine Americans aged 65 and older has Alzheimer's disease (AD), a proportion that increases to one in three among people 85 and older [1]. The aging of the U.S. population, as well as those in other industrialized countries, has resulted in marked growth in the numbers of older adults who live long enough to experience the debilitating impact of AD [2,3]. In addition to their increasing numbers, these older adults are also growing as a proportion of the total population. In the United States in 1900, there were about 3.1 million adults over the age of 65, and these individuals accounted for 4.1% of the total population. By contrast, in 2050, it is estimated that there will be 88.5 million adults over the age of 65, and this group will represent 20.2% of the population [4]. Assuming no new medical breakthroughs, it has been estimated that the number of AD cases will triple by 2050 from about 5 million to an estimated 13.8 million [5].

Advocacy organizations and policy makers have focused heavily on the need to develop effective treatments, service streams, and supports for AD. Passage of the 2011 National Alzheimer's Project Act (NAPA) [6] called for coordinated efforts to accelerate AD research, provide better care, and improve services for patients and families. NAPA also established an Advisory Council for Alzheimer's Research, Care, and Services. This group formulated a plan to address AD, including a clear set of objectives aimed at finding effective interventions and treatments [7].

Corinna E. Lathan[1*], Ian Coffman[1], Rita Shewbridge[1], Marissa Lee[1], Rosanna Cirio[2], Pasquale Fonzetti[2], P. Murali Doraiswamy[3] and Helaine E. Resnick[1]

[1]*AnthroTronix, 8737 Colesville Road, Suite L-203, Silver Spring, Maryland, USA*
[2]*Burke Rehabilitation Hospital, White Plains, New York, USA*
[3]*Duke Institute for Brain Sciences, Durham, North Carolina, USA*

*Address for Correspondence

Corinna E. Lathan, AnthroTronix, 8737 Colesville Road, Suite L-203, Silver Spring, Maryland, USA, Tel: 301-495-0770; Fax: 301-585-9075; E-mail: clathan@atinc.com

Among the objectives outlined by the NAPA Advisory Council are several that focus on caregivers. The vast majority of day-to-day care for people with AD is provided by informal caregivers, and the extent of this care is considerable. In 2013, Americans provided 17.7 billion hours of unpaid care to people with AD and other dementias, [8] and in 2014, more than 15 million family members and other unpaid caregivers provided care to these individuals [9]. This translates to 21.9 hours of care per caregiver each week, or 1,139 hours of care per caregiver each year. It is well-established that the vast majority of Alzheimer's care is provided in the home by unpaid caregivers.

The important role that is played by informal AD caregivers has generated growing interest in the characteristics and well-being of this population. National survey data show that 60% of caregivers of people with AD or dementia are adult children of the care recipient, 21% are over the age of 65, 51% are caring for someone over the age of 85, 23% have cared for the recipient for more than 5 years, 26% reported they have a disability, and nearly 17% report providing more than 40 hours of care each week. Ninety-four percent of AD caregivers in this survey reported that their care recipient experienced a change in thinking or memory in the past year [10].

Given the extent of their caregiving responsibilities, it is perhaps not surprising that AD caregivers are at increased risk of impaired cognition, depression, anxiety, and absenteeism, that they use healthcare services at higher rates than non-caregivers, and that their mental and physical health decreases as the severity of the AD care recipient's symptoms increases [11,12]. Numerous reports have shown that caregiving itself is associated with unfavorable effects on various aspects of cognitive function due to factors such as stress and depression [13-20].

The availability of convenient tools to assess cognitive

Table 1: Description of DANA subtests.

Test Name	Task Description
Simple Reaction Time (SRT)	The subject taps on the location of the yellow target symbol as quickly as possible each time it appears.
Procedural Reaction Time (PRT)	The screen displays one of four numbers (1,2,3 or 4) for 2 seconds. The subject taps the left button ("2" or "3") or right button ("3" or "4") at the bottom of the screen as quickly as possible to indicate which number was displayed.
Go/No-Go (GNG)	A house is presented on the screen with several windows. Either a "friend" (green) or "foe" (gray) appears in a window. The respondent must tap the "fire" button only when a "foe" appears.
Code Substitution-Learning (CSL)	Subjects refer to a code set of 9 symbol-digit pairs that are shown across the upper portion of the screen. Single symbol-digit pairs are presented in succession below the key, and the subject indicates whether or not the single pair matches the code by tapping "Yes" or "No."
Code Substitution-Recall (CSR)	After a delay of several intervening tests, the same symbol-digit pairs from the earlier Code Substitution-Learning task are presented without the code. The subject indicates whether or not the pairing was included in the code that was presented in the earlier code substitution learning section.
Spatial Processing (SP)	Pairs of four-bar histograms are displayed on the screen simultaneously, and the subject is requested to determine whether they are identical. One histogram is always rotated either ±90 degrees with respect to the other histogram.
Matching to Sample (MTS)	A single 4 x 4 checkerboard pattern is presented on the screen for brief study period. It then disappears for 5 seconds, after which two patterns are presented side-by-side. The subject indicates which of these two patterns was displayed during the study period.

Table 2: Successfully completed administrations (i.e.> 66% correct) ✓ and unsuccessfully completed administrations ✗ for subtest by participant.

Dyad	Participant Type*	Age	Gender	SRT1	PRT	GNG	CSL	CSR	SP	MTS	SRT2
1	Caregiver	74	Female	✓	✓	✓	✓	✗	✓	✗	✓
	Patient	73	Male	✓	✓	✓	✗	✗	✗	✗	✓
2	Caregiver	63	Female	✓	✓	✗	✓	✓	✓	✓	✓
	Patient	89	Male	✓	✗	✗	✗	✗	✗	✗	✓
3	Caregiver	77	Female	✓	✓	✓	✗	✗	✓	✓	✓
	Patient	90	Female	✓	✓	✓	✗	✗	✓	✓	✓
4	Caregiver	59	Female	✓	✓	✓	✓	✓	✓	✓	✓
	Patient	84	Female	✓	✗	✓	✗	✗	✗	✗	✓
5	Caregiver	74	Female	✗	✓	✓	✓	✗	✓	✓	✓
	Patient	81	Male	✓	✓	✓	✓	✗	✓	✗	✓
6	Caregiver	64	Female	✓	✓	✓	✓	✗	✓	✓	✓
	Patient	75	Male	✓	✓	✓	✗	✗	✗	✓	✓
7	Caregiver	70	Female	✓	✓	✓	✓	✓	✓	✓	✓
	Patient	75	Male	✓	✓	✓	✗	✗	✓	✓	✓

*All Caregivers were spouses except Caregiver 3 (friend) and Caregiver 4 (daughter).

performance is therefore applicable not only to AD patients but also to the caregivers themselves as a means to monitor their own cognitive trajectories. Ideally, such a tool would be easy to use, acceptable to both patients and caregivers, suitable for use in the home, and would provide real-time, actionable information that is useful to the caregiver for both caregiving and self-care. If successfully implemented, such a tool could help caregivers (1) by providing them with objective information to track the cognitive trajectories of AD care recipients, and (2) by providing information on their own cognitive performance so that they can better understand and respond appropriately to the challenges imposed by their caregiving role. However, translation of cognitive assessment tools from clinic-based to in-home use among Alzheimer's disease-caregiver dyads needs to be demonstrated not only to ensure that appropriate tests are selected for home use but also to show that these tests are sensitive to cognitive deficits as measured in the home, as this is where most

caregiving occurs.

With these considerations in mind, the objectives of this report are to: (1) assess the in-clinic feasibility of administering a battery of tests via a mobile cognitive performance instrument among Alzheimer's disease-caregiver dyads; (2) assess the sensitivity of this instrument for detecting mild cognitive impairment (MCI) and dementia and (3) test the feasibility of this instrument for assessing in-home cognitive performance.

Materials and Methods

Participants

AD patient-caregiver dyads were recruited at the Burke Rehabilitation Hospital in White Plains, New York. The Burke Rehabilitation Hospital is an acute rehabilitation hospital that provides inpatient and outpatient care services. AD Patients were
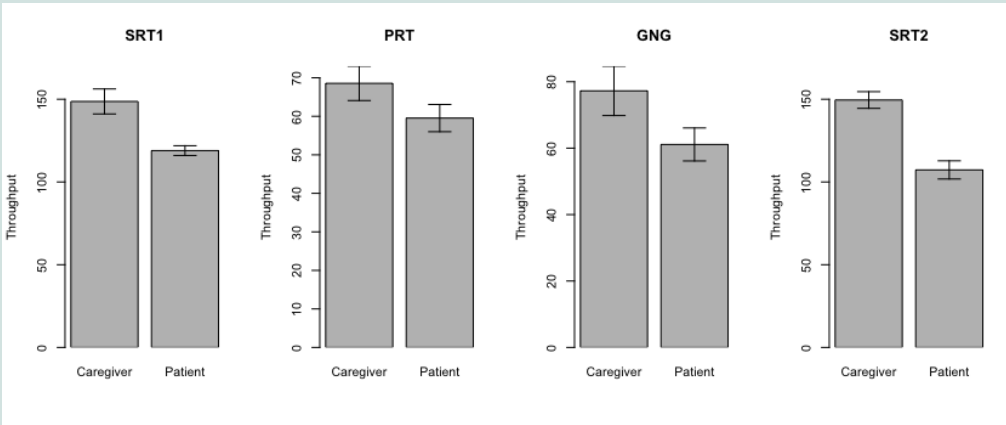
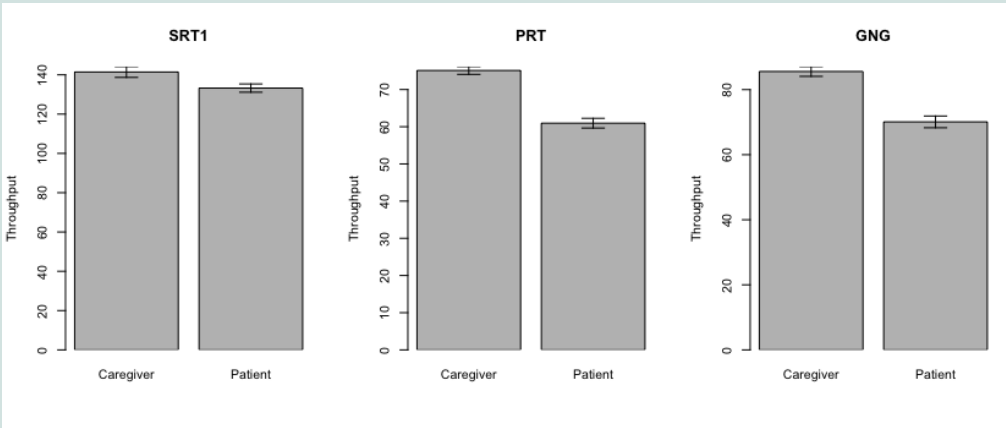**Figure 1:** Throughput for successfully completed subtests in-clinic. Error bars represent +/-1 standard error.

Figure 2: Throughput for successfully completed subtests in-home. Error bars represent +/- 1 standard error.
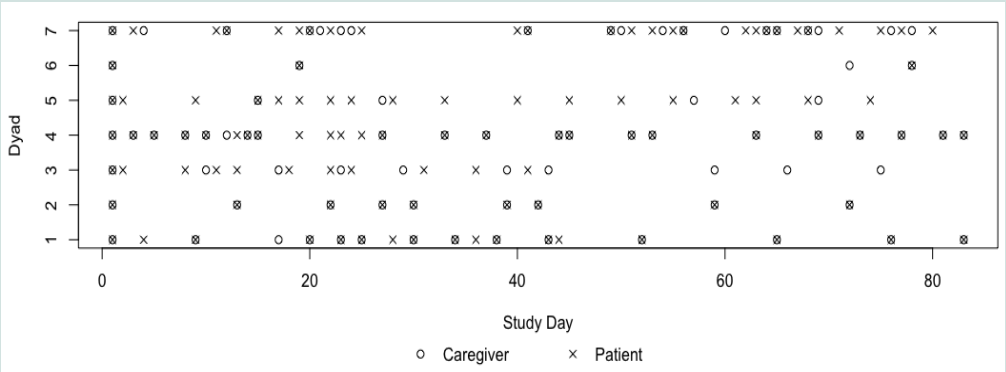

Figure 3: DANA administrations taken over the course of the study by dyad and group.

recruited from the outpatient Memory Evaluation and Treatment Service (METS) program, where patients are assessed and treated for memory disorders. Participants included patients diagnosed with mild Alzheimer's disease and their informal caregivers. The study was approved by the Institutional Review Board of Burke Rehabilitation Hospital.

Inclusion criteria for the dyads included minimum education and age requirements, a Geriatric Depression Scale score of less than six, and English language fluency. Caregiver-specific inclusion criteria also required no abnormal memory complaints, scores within normal range on the Mini Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) and no clinical diagnosis of dementia, mild cognitive impairment or Alzheimer's disease. Patient-specific inclusion criteria also required either no medications or stable history of medication usage for three months, meeting

NINCDS/ADRDA criteria for probable Alzheimer's disease and an MMSE score of greater than or equal to 20.

Prior to testing, patients and caregivers were screened by the site PI. Mild AD patients were established patients with previous diagnoses. The PI verified diagnoses and no new diagnostic screenings were conducted for the study. Caregivers were given standard neuropsychological tests (i.e. MMSE, MoCA). The site PI performed a history and neuropsychiatric exam to verify eligibility. Demographic information for the dyads is shown in Table 2.

Testing

All participants were administered DANA, a tablet-based, FDA-cleared neurocognitive assessment tool. DANA contains a battery of tests that is designed to examine cognitive performance on a number of distinct tasks, and its favorable psychometric properties and test-retest reliability have been documented [21,22]. A summary of the tests used in this study is provided in Table 1.

The primary outcome variable for each test is throughput (TP), a measure of cognitive efficiency.

Throughput relates speed and accuracy by quantifying the number of correct responses per minute:

$$TP = accuracy \times speed \times 60,000$$

where accuracy is the proportion of correctly completed trials, speed is the reciprocal of mean correct response time measured in milliseconds. The scaling factor of 60,000 converts the quantity to units of min^{-1}.

If a participant scored less than 66% correct on any test in the battery, results of that test were considered invalid and excluded from analysis. In the context of this study, such performance is indicative of the inability to perform a particular task.

Testing was carried out in two settings: a clinic-based setting at the Burke Rehabilitation Hospital and in patients' homes. The first testing session took place in clinic, where both caregivers and patients were administered the complete range of tests described in Table 1.[1] For in-home testing, each patient-caregiver dyad was provided with a tablet running DANA software and instructed to complete the assessment at home at least once a week for 90 days. For in-home testing, a complete administration consisted of the Simple Reaction Time, Procedural Reaction Time, and Go/No-Go subtests.

At the end of the 90-day home testing portion of the study, caregivers were contacted to take a follow-up survey soliciting feedback regarding their experience with DANA.

Results

In-clinic test administrations are shown in Table 2. AD patients were unable to reliably complete many of the tests that have been used previously in the DANA cognitive test battery, including CSL (1/7 patients completed), CSR (0/7), SP (3/7), and MTS (3/7). However, the patients had greater success in completing the simpler processing speed tasks: SRT1 (7/7), PRT (5/7), GNG (6/7), and SRT2

---

[1]The Simple Reaction Time subtest was administered twice: once at the beginning of the battery and once at the end. These two administrations are labeled as SRT1 and SRT2, respectively, in what follows.

(7/7). As indicated, caregivers were also unable to complete many of the assessments.

Figure 1 shows results for the four in-clinic tasks that were reliably completed. Two-sample Welch t-tests were used to assess differences between caregivers and patients for these subtests: SRT1 mean difference: 29.67 min^{-1}, t(6.52) = -3.66, 95% CI: -49.18, -10.22; PRT mean difference: 9.02 min^{-1}, t(9.99) = -1.58, 95% CI: -21.71, 3.68; GNG mean difference: 16.14 min^{-1}, t(8.72) = -1.81, 95% CI: -36.47, 4.19; SRT2 mean difference: 42.25 min^{-1}, t(11.91) = -5.66, 95% CI: -58.51, -25.98. Notice that group differences for the SRT1 and SRT2 subtests are significant at the 0.05 level.

The in-home phase of the study consisted of the SRT (single administration), PRT and GNG subtests. For these tests, both patients and caregiver performed similarly to in-clinic (Figure 2). Figure 3 shows the DANA administrations taken over the course of the in-home study by dyad and caregiver/patient group. Given the repeated measures aspect of the in-home administrations (i.e., multiple administrations nested under subject), multilevel regression models with intercepts estimated for each subject ID were used to evaluate the effect of Alzheimer's disease on throughput. The estimated effect was negative for all subtests (PRT: b = -12.80, 95% CI: -23.02, -2.57; GNG: b = -15.76, 95% CI: -29.35, -2.18; SRT: b = -16.22, 95% CI: -39.60, 6.93). Note that for the in-home phase, SRT was the only subtest not to reach significance at the 0.05 level.

Post-study follow-up interviews indicated that a majority of caregivers were able to independently set up the tablet and support the patient during the data collection period. Caregivers provided feedback on the device being used (a tablet) and the perceived usefulness of the in-home cognitive assessment. Caregivers provided additional feedback on DANA regarding instructions, stimulus size, and software navigation. Additionally, they reported generally positive impressions concerning perceived benefits of taking the assessment at home for both themselves and the patient.

Discussion

This study had three goals: (1) to assess the in-clinic feasibility of administering a battery of tests via a mobile cognitive performance instrument among Alzheimer's disease-caregiver dyads, (2) to assess the sensitivity of this instrument for detecting mild cognitive impairment (MCI) and dementia, and (3) to test the feasibility of this instrument for assessing in-home cognitive performance. Each is discussed below.

We found that DANA's full cognitive battery was not appropriate for our sample, particularly among AD patients. Patients were unable to reliably complete certain tasks such as Code Substitution, Matching to Sample, and Spatial Processing. Caregivers also had difficulty with some tasks, perhaps as a consequence of their advanced age. These tasks could potentially be modified for clinic use (such as increasing available response time). By contrast, simpler tasks like Simple Reaction Time, Procedural (Choice) Reaction Time, and Go/No-Go were generally reliably completed by both groups.

Despite our small sample size, in-clinic testing revealed numerical trends consistent with the expected result that the Alzheimer's group would perform worse than caregivers across a range of cognitive

tests. These trends were also observed in the home, suggesting a reliable transfer of DANA's sensitivity to Alzheimer's disease that was demonstrated in the clinical setting. Although in some cases differences in cognitive performance between caregiver and patient groups failed to reach traditional significance thresholds, we believe that more consistent results will be obtained with larger sample sizes and/or through measurement of factors that are likely to contribute to variance in cognitive performance (e.g., medication, stress, etc.).

Finally, our results speak to the feasibility of using a portable neurocognitive assessment tool in the home. Although the number of administrations varied among participants, testing sessions spanned the entire range of the study period and were generally evenly distributed across it (Figure 3), suggesting that engagement with the device was consistent over the course of the study. Results of the follow-up questionnaire provided useful insights into the usability concerns among caregivers and patients, thereby providing a platform for further development of this testing modality in this population.

An important element of our findings relates to identification of strategies that simultaneously enhance both patient- and caregiver-centered support among people whose lives are affected by AD. For caregivers, one aspect of these strategies involves providing self-care tools that help them assess cognitive performance in a manner that optimizes their ability to care for themselves and the people who depend on them [23]. Availability of these caregiver-centered tools is tied to the economic value of informal caregiving. A recent report indicated that informal dementia caregiving is valued at $218 billion annually [7]. Because there is no resource available to cover the cost of replacing informal dementia care with paid support, efforts to ensure caregivers' well-being including their ability to care for people with AD and to care for themselves have clear economic and policy implications for countries whose populations continue to age without any obvious service streams to support these demographic changes.

Our findings can also be interpreted in the context of NAPA's plan to address AD. Goal 3 of that plan is to "Expand support for people with Alzheimer's disease and their families." Strategy 3B of that plan calls for enabling "family caregivers to continue to provide care while maintaining their own health and well-being" and strategy 3C seeks to "Assist families in planning for future care needs" [24]. Given that about half of all AD caregivers are themselves over the age of 55 - a finding that is reflected in our data - the needs of these aging caregivers must be addressed in parallel with the needs of their care recipients. Our data on the feasibility of using a home-based cognitive assessment tool is consistent with federal priorities to support caregivers as well as AD patients with tools that support their needs, help maintain caregiver health, and assist in planning for future care needs.

It is the context of supporting caregivers in this important role that our findings are especially relevant in a public policy context. The average per-person Medicare spending for seniors with Alzheimer's is almost three times higher than average per-person spending for all other seniors. Under Medicaid, spending is 19 times higher [24]. It is important to stress that these costs are associated with formal health care provision, and that effective provision of informal care helps to keep these costs down.

Although the aging of the population will undoubtedly result in increasing numbers of older adults who will continue to incur substantial costs to the formal health care system, it may be possible to control these costs by offering caregivers effective tools that can optimize their ability to provide informal care.

References

1. Hebert LE, Weuve J, Scherr PA, Evans DA (2013) Alzheimer's disease in the United States (2010-2050) estimated using the 2010 Census. Neurology 80: 1778-1783.

2. Ortman JM, Velkoff VA, Hogan H (2014) An aging nation: the older population in the United States: Current population reports. US Census Bureau, Washington, DC, pp. 25-1140.

3. Eurostat Statistics Explained (2016) Population structure and ageing.

4. ACL (2016) Administration on Aging (AoA) projected future growth of the older population.

5. Lathan C, Spira JL, Bleiberg J, Vice J, Tsao JW (2013) Defense Automated Neurobehavioral Assessment (DANA) - psychometric properties of a new field-deployable neurocognitive assessment tool. Mil Med 178: 365-371.

6. Authenticated US Government Information (2011) National alzheimer's project act. pp. 111-375.

7. NASPE (2015) National plan to address alzheimer's disease: 2015 update. pp. 1-58.

8. Alzheimer's Association (2014) 2014 Alzheimer's disease facts and figures. Alzheimers Dement 10: e47-e92.

9. Alzheimer's Association (2015) 2015 Alzheimer's disease facts and figures. Alzheimers Dement 11: 332-384.

10. Caregiving Across the United States (2009) Caregivers of persons with alzheimer's disease or dementia in 8 states and the district of Columbia.

11. Kannan H, Bolge SC, Del Valle M, Alvir J, Petrie CD (2011) The association between Alzheimer's disease symptom severity and caregiver outcomes: a cross-sectional study. Prim Care Companion CNS Disord 13.

12. Pinquart M, Sorensen S (2003) Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 18: 250-267.

13. Lathan C, Wallace AS, Shewbridge R, Ng N, Morrison G, et al. (2016) Cognitive health assessment and establishment of a virtual cohort of dementia caregivers. Dement Geriatr Cogn Dis Extra 6: 98-107.

14. Mackenzie CS, Smith MC, Hasher L, Leach L, Behl P (2007) Cognitive functioning under stress: evidence from informal caregivers of palliative patients. J Palliat Med 10: 749-758.

15. Caswell LW, Vitaliano PP, Croyle KL, Scanlan JM, Zhang J, et al. (2003) Negative associations of chronic stress and cognitive performance in older adult spouse caregivers. Exp Aging Res 29: 303-318.

16. Lee S, Kawachi I, Grodstein F (2004) Does caregiving stress affect cognitive function in older women? J Nerv Ment Dis 192: 51-57.

17. Oken BS, Fonareva I, Wahbeh H (2011) Stress-related cognitive dysfunction in dementia caregivers. J Geriatr Psychiatry Neurol 24: 191-198.

18. Gaugler JE, Davey A, Pearlin LI, Zarit SH (2000) Modeling caregiver adaptation over time: the longitudinal impact of behavior problems. Psychol Aging 15: 437-450.

19. Vitaliano PP, Zhang J, Young HM, Caswell LW, Scanlan JM et al. (2009) Depressed mood mediates decline in cognitive processing speed in caregivers. Gerontologist 49: 12-22.

20. Corrêa MS, Vedovelli K, Giacobbo BL, de Souza CE, Ferrari P, et al. (2015) Psychophysiological correlates of cognitive deficits in family caregivers of patients with Alzheimer Disease. Neuroscience 286: 371-382.

21. Russo CR, Lathan CE (2015) An evaluation of the consistency and reliability

of the defense automated neurocognitive assessment tool. Appl Psychol Meas 39: 566-572.

22. Miller LS, Lewis MS, Williamson GM, Lance CE, Dooley WK, et al. (2006) Caregiver cognitive status and potentially harmful caregiver behavior. Aging Ment Health 10: 125-133.

23. Chari AV, Engberg J, Ray KN, Mehrotra A (2015) The opportunity costs of informal elder-care in the United States: new estimates from the American Time Use Survey. Health Serv Res 50: 871-882.

24. Alzheimer's Association (2016) Costs of alzheimer's to medicare and medicaid.

From battlefield to home: a mobile platform for assessing brain health

Helaine E. Resnick, Corinna E. Lathan

AnthroTronix Inc., Silver Spring, MD, USA

Correspondence to: Corinna E. Lathan, PhD. AnthroTronix, 8737 Colesville Road, Suite L-203, Silver Spring, MD 20910, USA. Email: clathan@atinc.com.

Abstract: Cognitive testing batteries have been used for decades to diagnose deficits associated with conditions such as head injury, age-related cognitive decline, and stroke, and they have also been used extensively for educational evaluation and planning. Cognitive testing is generally office-based, administered by professionals, uses paper and pencil testing modalities, reports results as summary scores, and is a "one shot deal" whose primary objective is to identify the presence and severity of cognitive deficit. This paper explores innovative departures from historical cognitive testing strategies and paradigms. The report explores (I) a shift from disease diagnosis in the office setting to mobile tracking of cognitive health and wellness in any setting; (II) the strength of computer-based cognitive measures and their role in facilitating development of new computational methods; and (III) using cognitive testing to inform on individual-level outcomes over time rather than dichotomous metrics at a single point in time.

Keywords: Cognitive testing; computational methods; cognition; neuroinformatics

Introduction and overview

In addition to regulating breathing, heart rate, and blood pressure, the brain receives information from the environment, interprets this information, and guides appropriate responses to these stimuli. From an evolutionary point of view, an organism's ability to effectively process external information is advantageous because it facilitates survival (1). In prehistoric times, the ability to react quickly to visual stimuli could make the difference between a successful hunt and starvation. In the modern era, efficient processing of external information has implications for tasks as diverse as identifying the best moment to swing a baseball bat, being able to distinguish friend from foe on the battlefield, and being able to recognize and respond to traffic signals. Cognitive efficiency refers to how quickly and accurately one can process information, and this aspect of brain function has far-reaching implications for well-being throughout the life span and into old age (2-4). The importance of assessing and maintaining cognitive efficiency

has led to development of tools to measure various aspects of brain health and function (5).

Historically, cognitive testing has been conducted in office-based settings by specially-trained professionals such as neuropsychologists. Cognitive batteries that are administered in these settings include intelligence tests, finger tapping tests, trail making tests, coding tests, letter-number sequencing tests, verbal learning tasks, and block design tests. These tests evaluate different aspects of brain function, including cognitive efficiency or processing speed, spatial processing, visual scanning and attention, immediate recall, short-term memory, working memory, language, attention/concentration, executive function, and visual-spatial discrimination (6). The variety and complexity of brain functions that are assessed by cognitive testing batteries hint at the many ways that deficits in these functions can unfavorably impact daily function. For example, an injured student athlete's grades may decline, a cognitively impaired older adult may lose keys or leave the stove on, and an injured soldier may put himself and his

unit at risk.

The emergence of mHealth offers an opportunity for radical changes in how we assess cognitive or brain health, and this report explores four considerations related to this paradigm shift: (I) limitations of traditional approaches to cognitive testing; (II) opportunities for mobile assessment of brain health; (III) mobile platforms for patient-centered cognitive assessment; and (IV) re-thinking data and outcomes. These considerations reveal three broad themes related to the evolution of cognitive efficiency testing: A shift from disease diagnosis in the office setting to mobile tracking of health and wellness in any setting; the strength of computer-based measures and their role in facilitating development of new computational methods, and the use of cognitive testing to inform on individual-level outcomes over time rather than dichotomous metrics at a single point in time.

Limitations of traditional approaches to cognitive testing

By definition, identification of a cognitive deficit is required before appropriate interventions can be implemented. For this purpose, traditional paper and pencil testing is reliable, valid, and has diagnostic value—all important features for clinical application. However, traditional approaches to assessment of brain health also come with a number of important limitations. These tests are time-intensive for both testers and patients; special training and testing areas are needed; they are expensive; it can be difficult to get short-term evaluations because access to neuropsychological services is limited; there are learning effects that can't be mitigated by alternate forms of the tests, and the tests were not designed to be patient-centered tools or to assess how people function in community-based settings (7,8).

In addition to these issues, there are important limitations related to the nature of the data, how they are collected, and how these factors interact to impact usability (9). For example, because paper and pencil tests are not computerized, factors related to how these tests are administered can impact scoring across testing environments. The tests do not permit export of raw or summary data in a manner that facilitates data analysis or integration with patients' electronic health records. Test batteries frequently yield simple summary scores on various sub-tests, a system that does not offer insight into complex response patterns that may provide important insight on the presence or origin of various aspects of cognitive

deficit. Finally, these testing modalities focus heavily on data collection at a single point in time, and comparisons of these cross-sectional measures to population-based norms. Thus, they are not designed to track individuals' cognitive efficiency over time, nor are they designed to put patient data in patients' hands where this information can be acted upon when a meaningful change in performance occurs.

Opportunities for mobile assessment of brain health

In recent years there has been a call for development and broad implementation of computerized cognitive testing. This need has been highlighted by stakeholders including drug developers, federal agencies that sponsor research focused on cognitive outcomes, and from clinicians who wish to move toward testing strategies that provide greater access to cognitive data in a manner that offers faster, more detailed information without sacrificing quality or increasing patient burden (10,11). In addition to these stakeholders, patients and caregivers are also developing higher expectations concerning the quality of, and access to their own health-related data (12,13). Mobile cognitive testing responds to stakeholder demands, offering a number of advantages over traditional methods, including considerations related to ease of administration and access to data.

Beyond these obvious advantages, mobile cognitive testing is patient-centered, allowing patients unprecedented access and insight into their own cognitive efficiency at a single point in time as well as understanding of patterns of change over time. The value of this information is not limited to patients. The vast majority of care for people with chronic disease comes from informal caregivers—most often from adult children and elderly spouses (14). The availability of a mobile platform that caregivers can use to assess a care recipient's cognitive efficiency may offer new opportunities for caregivers to reliably track cognitive change over time, thereby enabling them to be more effective caregivers. Meeting these needs is consistent with federal priorities concerning the need to help "family caregivers to continue to provide care while maintaining their own health and well-being." (15).

The variety of settings in which cognitive deficits can impact day to day function—the baseball field, the battle field, nursing homes, and community-based residences—reflect the value of having mobile cognitive assessment tools that can be used effectively in diverse settings. These

technologies can also be used repeatedly over time in a manner that informs on clinically meaningful trends, and that puts actionable information directly in the hands of consumers.

Finally, the limitations of traditional cognitive testing highlight the ways in which a new generation of mobile cognitive efficiency testing strategies can meet evolving patient needs. For example, assessment of cognitive efficiency at during primary care visits would establish individual baseline, allowing highly sensitive assessments of changes that might occur due to a sports injury, depression, or age-related dementia. Mobile tracking could also enable measurement-based care. Underscoring the desire to base healthcare on objectives measures, the Kennedy Forum recently issued a national call to expand the practice of measurement-based care from medical and surgical fields to behavioral health (16).

Mobile platforms for patient-centered cognitive assessment

Advances in technology, improved health literacy, and the independence that has been fostered by mobile technologies have all contributed to a shift in patients' expectations of their interactions with the healthcare system and their own health information. Patients— particularly younger patients—expect to access their health data and health care providers in ways that were unthinkable 15 years ago. Many primary care practices offer portals that allow patients to make appointments online, to access their laboratory results, and to request prescriptions refills. Policy-driven incentives encouraging primary care providers to adopt electronic health records, combined with ongoing efforts to enhance patient access with mobile technology reflect broader trends that recognize the importance of technology-enhanced patient-centered care (17). Patients have developed a new set of expectations concerning access to their own health data, and there is a new sense of autonomy among patients that reflects the desire to have a greater degree of data-driven control over their health and wellness (18).

It is against this backdrop that traditional strategies to assess cognitive performance should be re-evaluated. If cognitive testing can be conducted reliably outside of an office setting, it is reasonable to expect that these testing strategies should be taken into the field—taken to patients— rather than continuing to expect patients to come to the office. A key principle of "patient centered care" is the idea

that patients are the best source of information about how well their health care providers are meeting their needs, and those patient perceptions about their healthcare delivery correlate with both health outcomes and satisfaction with care (19).

Although age-related cognitive decline is not the only setting in which mobile brain health technologies provide benefit to patients and families, this setting provides a useful framework to think about the value of these strategies. Among older adults, there are numerous non-office settings where cognitive testing could provide useful information to both formal and informal caregivers. These have direct application to patient-centered care because it is well-established, for example, that older adults have a strong preference to remain independent in their homes as long as possible. Such preferences, along with the recognized cost advantages of providing community-based—as opposed to institutional—care for frail seniors, is at the root of a shift toward development of systems for community-based provision of long term care supports and services. A key element of care plans that are implemented in the community is a clear understanding of care recipients' cognitive status. The frailty of this population, along with a focus on home-based care reflect the value of mobile assessment tools that can provide integrated care teams with information on cognitive status over extended periods of time in a manner that informs on diverse aspects of care for growing numbers of seniors.

The value of this information can be interpreted in the context of the diversity of settings in which older adults reside, and in which tracking of their cognitive efficiency would be useful not only to them, but also to both formal and informal caregivers. For older adults who use nursing home services, ongoing assessment of cognitive efficiency could inform directly on various aspects of institutional care, and this information could be readily collected in this care setting using mobile platforms, and it would be available not only clinicians, but also to patients and family members. Intermediate between community/home-based residential settings and institutional care are assisted living settings in which seniors receive a limited set of health services. Like nursing home settings, care teams in these residential settings could benefit from easy access to reliable data on seniors' cognitive efficiency. The ability of mobile technologies to dovetail with electronic health records would further enhance continuity of care for frail older adults who receive care from numerous specialists who practice in these diverse care settings.

Re-thinking data and outcomes

In recent years, there has been a tremendous increase in awareness of the role that "big data" can play in clinical decision-making, including how it can be used to personalize cognitive health (20,21). There is parallel interest in the idea that objective data should be at the foundation of individualized decisions about health, and that generation of, and access to clinical data should extend beyond the doctor's office; it should be tailored to the needs of individual patients, it should provide insight on patients' longitudinal health trends, the information should available to patients and their families on demand, and data should be available using technologies that are chosen by consumers (22).

Among the drivers of the increased emphasis on collection of individualized data for cognitive assessment in particular is the aging of the U.S. population, sometimes called "the graying of America" or the "silver tsunami". Growing numbers of older adults have resulted in a marked increase in the burden of Alzheimer's disease and other dementias. These burdens not only impact patients, but they also have unfavorable effects on informal caregivers as well as the formal healthcare system.

New mobile technologies capture, export, and facilitate analysis of computerized cognitive data in a manner that enables use of all data that are collected by these technologies, not just summary scores. Why is this important for cognitive testing? Unlike many biological determinations (e.g., blood glucose or cholesterol) where a single threshold measure can unambiguously define the presence or absence of a disease or risk state, cognitive deficits can be subtle, and they can occur in multiple areas of brain function. Assessment of the presence or absence of a cognitive health condition that requires intervention may require many tests that evaluate multiple brain functions, often using a single summary score that is supposed to capture a multitude of complex patterns and functions.

Historically, cognitive testing scores are collapsed so that cognitive status is presented as binary (impaired/not impaired) or ordinal (normal/mild impairment/moderate impairment/severe impairment). This framework has important limitations. It is constrained by the maxim values that are dictated by the sum of scores on component tests. Pooling of sub-scores can obscure profound cognitive impairment on one subtest while still showing a favorable overall score. A third limitation involves the assumption that a single overall score offers the greatest clinical utility

and that patterns of fluctuation over the course of many trials are of little or no value to cognitive assessment or care planning (23).

We believe that efforts to optimize the richness of computerized cognitive testing data must fully utilize all trial-by-trial data that are offered by these technologies because of the tremendous insight that this highly granular information can provide. These strategies offer an opportunity to depart from a traditional framework that relies on a single set of summary scores to one in which new computational methods can capitalize on many thousands of data points to provide insight on subtle changes in cognitive efficiency over time. The growing use of mobile cognitive assessment technologies will only enhance the impact of these efforts because of their ability to facilitate access to this information on the part of patients and families.

It is helpful to use a specific example to illustrate some of these concepts. Simple reaction time (SRT) assesses psychomotor speed—often in response to a visual stimulus, and the test often involves between 20 and 50 trials depending on the tool or instrument. Historically, SRT summary scores have been used as a means to describe an individual's global performance on this subtest at a single point in time without regard to quantifying the shape of the curve that is generated by performance on each trial, and without appreciable attention to how response patterns may change over time. We propose a new focus that uses all the data that are available from newer mobile technologies to provide both a more granular view of an individual's cognitive efficiency at a given point in time, and to help quantify meaningful changes over time.

This new focus can potentially unlock new applications of cognitive testing data in a quantitative and clinically relevant framework that is consistent with evolving expectations of patient-centered care. These methods could reveal previously-unidentified deficits and perhaps the etiology of some forms of cognitive dysfunction. An example of this strategy is presented in *Figure 1*. This figure shows SRT data from young adults at sea level, and the same adults at extreme altitude where their cognitive efficiency was greatly diminished due to hypoxia. These data, which were collected with a hand-held mobile cognitive assessment instrument, reveal significant differences in data patterns between the two groups, with hypoxic individuals' SRT data being significantly more unstable than their uninjured counterparts. Examination of simple means and standard deviations do very little to fully utilize the richness of these data. We continue to develop these and other new
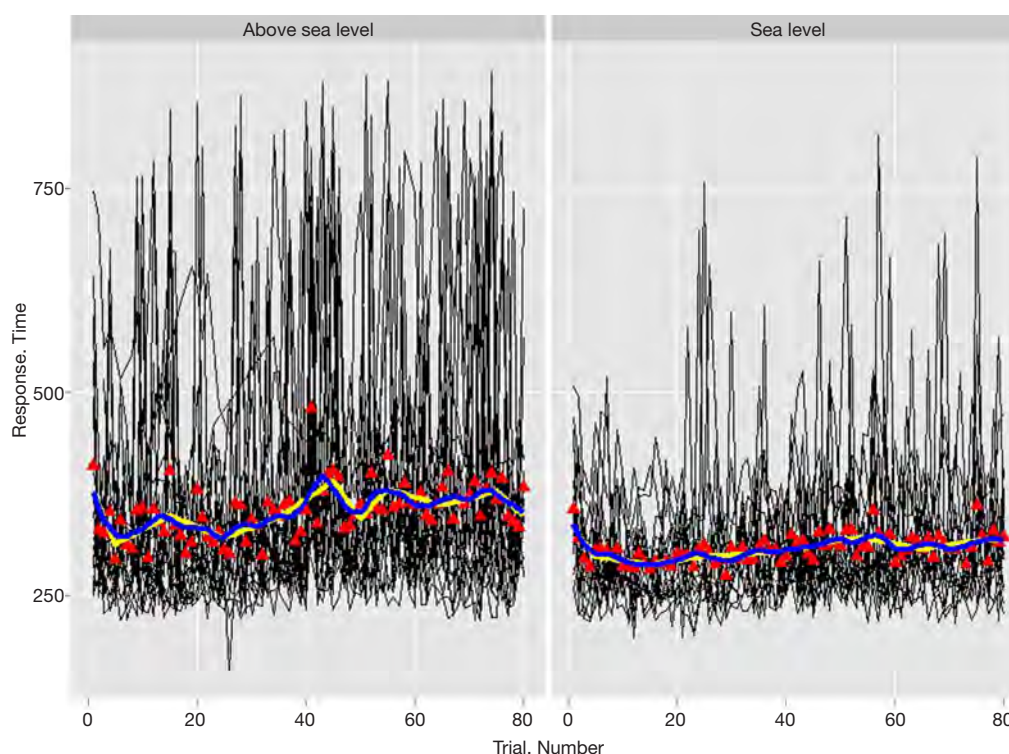
Figure 1 Trial-by-trial analysis of simple reaction time testing among subjects at sea level and at extreme altitude.

computational methods to meet the growing expectations of patients and caregivers who are coming to expect more than a simple "yes or no" concerning questions about their health status.

Conclusions

Mobile platforms for computerized cognitive testing offer new opportunities to put actionable health information in the hands of consumers, to develop novel computational strategies that fully leverage large amounts of highly detailed cognitive efficiency data, and to meet the needs of diverse populations in a fully patient-centered framework.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Survival – mind and brain. Available online: https://thepsychologist.bps.org.uk/volume-24/edition-1/survival-%E2%80%93-mind-and-brain (accessed June 3, 2016).
2. Hoffman B. Cognitive efficiency: A conceptual and methodological comparison. Learning and Instruction 2012;22:133-44.
3. Rypma B, Berger JS, Prabhakaran V, et al. Neural correlates of cognitive efficiency. Neuroimage 2006;33:969-79.
4. Derakshan N, Eysenck MW. Anxiety, processing efficiency, and cognitive performance new developments from attentional control theory. European Psychologist 2009;14:168-76.
5. Cullen B, O'Neill B, Evans JJ, et al. A review of screening tests for cognitive impairment. J Neurol Neurosurg Psychiatry 2007;78:790-9.
6. Harvey PD. Clinical applications of neuropsychological assessment. Dialogues Clin Neurosci 2012;14:91-9.
7. Sbordone RJ. Limitations of neuropsychological testing to predict the cognitive and behavioral functioning

of persons with brain injury in real-world settings. NeuroRehabilitation 2001;16:199-201.

8. Johnstone B, Frank RG. Neuropsychological assessment in rehabilitation: Current limitations and applications. NeuroRehabilitation 1995;5:75-86.

9. Wesnes KA. Moving beyond the pros and cons of automating cognitive testing in pathological aging and dementia: the case for equal opportunity. Alzheimers Res Ther 2014;6:58.

10. Wild K, Howieson D, Webbe F, et al. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement 2008;4:428-37.

11. Bauer RM, Iverson GL, Cernich AN, et al. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012;27:362-73.

12. Caine K, Tierney WM. Point and counterpoint: patient control of access to data in their electronic health records. J Gen Intern Med 2015;30 Suppl 1:S38-41.

13. Mandl KD, Kohane IS. Time for a patient-driven health information economy? N Engl J Med 2016;374:205-8.

14. 2015 Alzheimer's Disease Facts and Figures. Available online: https://www.alz.org/facts/downloads/facts_figures_2015.pdf (accessed June 3, 2016).

15. National plan to address Alzheimer's disease: 2015 update. Available online: https://aspe.hhs.gov/national-plan-address-alzheimers-disease-2015-update#strategy3 (accessed June 3, 2016).

16. A National Call for Measurement-Based Care. Available online: https://www.thekennedyforum.org/news/measurement-based-care-issue-brief (accessed June 13, 2016).

17. Health IT Legislation and Regulations. Available online: https://www.healthit.gov/policy-researchers-implementers/health-it-legislation-and-regulations (accessed June 3, 2016).

18. Available online: https://www.healthit.gov/patients-families/your-health-data (accessed June 3, 2016).

19. Epstein RM, Street RL Jr. The values and value of patient-centered care. Ann Fam Med 2011;9:100-3.

20. Geerts H, Dacks PA, Devanarayan V, et al. From big data to smart data in Alzheimer's disease. The brain health modeling initiative to foster actionable knowledge. Alzheimers Dement 2016. [Epub ahead of print].

21. Cheung MW, Jak S. Analyzing big data in psychology: a split/analyze/meta-analyze approach. Front Psychol 2016;7:738.

22. Maglaveras N, Kilintzis V, Koutkias V, et al. Integrated Care and Connected Health Approaches Leveraging Personalised Health through Big Data Analytics. Stud Health Technol Inform 2016;224:117-22.

23. Wang L, Zhang Z, McArdle JJ, et al. Investigating ceiling effects in longitudinal data analysis. Multivariate Behav Res 2009;43:476-496.

Trial-by-Trial Pattern Analysis: A Novel Strategy for Identifying Neurocognitive Deficit With Computerized Cognitive Tests

Darcy Steeg Morris, PhD,[1] Corinna Lathan, PhD,[2] Lisa Weissfeld, PhD,[3] Timothy Lacy, MD,[2] Helaine E. Resnick, PhD, MPH[1]

[1]Resnick, Chodorow and Associates, Silver Spring, MD, [2]Anthrotronix, Inc., Silver Spring, MD, Statistics Collaborative, Inc., Washington, DC[3]

Abstract

Background: Computerized cognitive testing evaluates numerous aspects of brain function. In many cases, these tests (e.g. simple reaction time, SRT) involve multiple trials that are collapsed into a summary performance score, often a mean. In turn, these summary scores—used alone, or in combination with scores from other subtests—are used to assess cognitive health both cross-sectionally, and over time. Aggregation of quantitative data from multiple trials into a summary mean for a given test assumes that (1) a mean provides an accurate reflection of an individual's performance across all trials, and (2) no clinically relevant information can be gleaned from the shape of an individual's response curve across the trials. **Methods:** We challenged these long-held assumptions by taking advantage of the richness of trial-by-trial data from computerized cognitive testing to develop a strategy to identify clinically distinct groups from each other based on the pattern of their responses, rather than the mean. Using SRT data as a test case, we applied this method to the settings of concussion and altitude-induced hypoxia with data from 6 concussed and 153 non-concussed Air Force Academy Cadets and data collected in 21 college-aged students who were tested at sea level and again at 5,260m. We first plotted individual-level responses across 40 SRT trials, followed by trial-specific means. We fit loess curves to these means, and then fit linear spline models with a random intercept to these curves

$$E(Response\ Time|Trial\ Number)$$
$$= \beta_0 + \beta_1 * Trial\ Number + \sum_{k=1}^{K} \beta_{1+k} * (Trial\ Number - \kappa_k)_+ + u_i$$

where i denotes the subject index and u_i is the random intercept. **Results:** Our results showed that at sea level, subjects' mean response times stabilized after a brief learning period in the first 5 trials, and an identical pattern was observed in non-concussed cadets. Under hypoxic conditions, subjects' mean SRT responses did not stabilize over the 40 trials, remaining erratic throughout the test, and this was also observed in concussed cadets. Loess curves confirmed localized curvature across trials among hypoxic and concussed subjects relative their non-concussed and sea level controls. Spline models were run separately for each data set, and these showed statistically different slopes at several points across the 40 trials for concussed vs. non-concussed cadets, and for sea level vs. hypoxic subjects. **Conclusions:** By leveraging the richness and nuance of trial-by-trial responses to computerized cognitive testing, our results offer promise for developing pattern-based screening and treatment monitoring tools. If further developed, these strategies could be applied in the settings of traumatic brain injury, concussion, depression, PTSD and other conditions where return to duty decisions can benefit from inexpensive, objective, and nuanced data on cognitive performance.

Methods

DANA is a hand-held, FDA-cleared clinical neurocognitive assessment tool that measures and tracks changes in cognitive efficiency by measuring response speed and accuracy. DANA includes eight cognitive tests and seven psychological questionnaires that measure multiple aspects of brain health. DANA has been validated in diverse military and civilian research settings. We report the evolution of our repeated measures work using DANA's trial-by-trial simple reaction time (SRT) data from three data sources:

Ft. Hood: 219 psychologically "healthy" and 98 "unhealthy" service members (i.e., CES > 8, PHQ > 9, PCL > 49)

Altitude data: 17 people at sea level and 21 at extreme altitude

Air Force Academy athletes: 153 normal and 6 concussed

Step 1: Visualize group means over time

Step 2: Fit Loess curves to visualize the shape of an ideal smoothed curve for "normal" and "non-normal" groups' repeated SRT measures.

Step 3: Use Loess curves as a "target" for spline regression modeling for "normal" and "non-normal" groups to confirm that there are different shapes.

Step 4: Use unsupervised, machine-based learning algorithms—group based trajectory modeling—to automatically cluster trajectories that are "similar" to each other from a statistical point of view.

Step 5: Assess how well these clusters identify "normal" from "non-normal" individuals.

Results

Step 1: Visualize group means over time
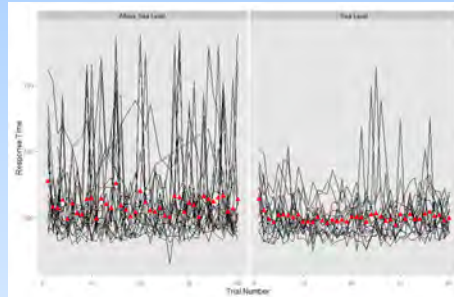
Interpretation: Young subjects who were at extreme elevation—5620 m—had more varied, group-level means across 40 SRT trials than the same individuals when they were at sea level. The red triangles show the mean SRT at each trial for the entire study group. This variability is not fully captured by examining summary means of the 40 trials.

Step 2: Fit Loess curves to visualize the shape of an ideal smoothed curve for "normal" and "non-normal" groups' repeated SRT measures.
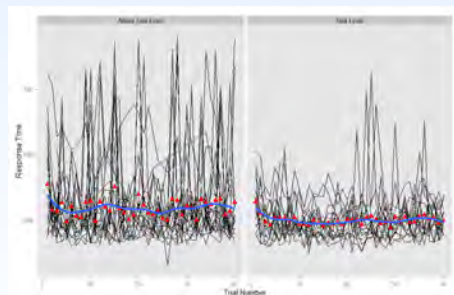
Interpretation: The blue loess curve is the "ideal" smoothed curve that fits the trial-by-trial group means that are shown by the red triangles. Any modeling technique that is used to analyze these data to test for group differences in the shape of the curves should approximate the smoothed loess curve that is shown in blue.

Results

Step 3: Use Loess curves as a "target" for spline regression modeling for "normal" and "non-normal" groups to confirm that there are different shapes.
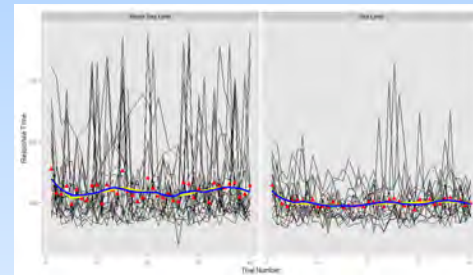
Interpretation: Linear spline regression analysis captures the shape of the average profiles and indicated statistically significant differences in the shape of the average profile for subjects at sea level and at extreme altitude.

Step 4: Use unsupervised machine learning techniques (longitudinal k-means clustering shown here) to uncover hidden clusters of "normal" subjects in the data.
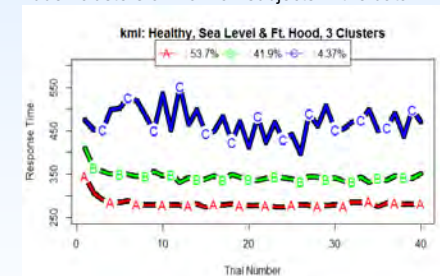
Interpretation: When the "normal" subjects are forced into three clusters, most subjects belong to Cluster A, with the fastest and least variable response times. Clusters B and C are characterized by slower and more variable response times, representing 41.9% and 4.37% of subjects, respectively.

Results

Step 5: Use the k-means results to predict group membership of each out-of-sample ("non-normal") subject by choosing the smallest Minkowski distance (Euclidian case) between their trial-by-trial trajectories and the centers of Clusters A, B and C:

$$d(y_i, \overline{y_k}) = \left(\sum_{t=1}^{T} |y_{it} - \overline{y_{kt}}|^2\right)^{1/2}$$

where y_is make up a vector of response times for a "non-normal" subject i, $\overline{y_k}$s are a vector of mean reaction times for all "normal" subjects in cluster k, and t is an index of trial number. Results:

	Group A	Group B	Group C
Above sea level	8 (38.1%)	11 (52.4%)	2 (9.5%)
Concussed	3 (50.0%)	2 (33.3%)	1 (16.7%)
Ft. Hood "unhealthy"	49 (50.0%)	45 (45.9%)	4 (4.1%)

Interpretation: In all cases, the proportion of "non-normal" subjects assigned to Cluster A is less than the proportion of assigned "normal" subjects. Accordingly, more "non-normal" subjects are classified as belonging to Clusters B and C relative to the "normal" subjects. This means that "non-normal" subjects' trajectories tend to be more similar to either Clusters B or C, which are characterized by slower and more variable response times relative to Group A, to which most "normal" subjects belong.

Conclusion/Future directions

By leveraging the richness and nuance of trial-by-trial responses to computerized cognitive testing, our results offer promise for developing pattern-based screening and treatment monitoring tools. If further developed, these strategies could be applied in the settings of traumatic brain injury, concussion, depression, PTSD and other conditions where return to duty decisions can benefit from inexpensive, objective, and nuanced data on cognitive performance.

While these initial results are promising, there are some limitations that will need to be addressed in future work. For example, k-means clustering does not allow for the inclusion of covariates (e.g., age, gender, etc) that may be predictive of group membership. However, we are actively working with other machine learning techniques, such as group-based trajectory modeling, that can address this issue.